U.S. Department
of Transportation

**Federal Highway
Administration**

Publication No. FHWA-RD-91-124
March 1993

DEVELOPMENT OF RELATIONSHIP BETWEEN TRUCK ACCIDENTS AND GEOMETRIC DESIGN: PHASE I

Office of Research and Development
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, Virginia 22101-2296

FOREWORD

This report, FHWA-RD-91-124, contains research findings on the development of statistical relationships between truck accidents and key highway geometric design elements. The results in this report will be of interest to those concerned with highway design and operations, as well as those involved in developing and establishing large-truck safety regulations.

First, statistical frameworks suitable for describing such relationships were proposed. Preliminary models were then developed using existing accidents and road inventory data for three roadway classes: rural Interstate, urban Interstate and freeways, and rural two-lane undivided arterial. The performance of the developed models was statistically evaluated, and a "data needs" study was conducted to identify additional data and variables for enhancing the predictive capability of the preliminary models and for developing representative "National" models in the future.

Sufficient copies of the report are being distributed to provide a minimum of one copy to each FHWA regional and division office, and three copies for each State highway agency. Direct distribution is being made to the division offices. Additional copies for the public are available from the National Technical Information Service (NTIS), Department of Commerce, 5285 Port Royal Road, Springfield, Virginia 22161. A small charge will be imposed by NTIS.

Lyle Saxton, Director
Office of Safety and Traffic
Operations Research and Development

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-RD-91-124		

4. Title and Subtitle	5. Report Date August 1993
DEVELOPMENT OF RELATIONSHIP BETWEEN TRUCK ACCIDENTS AND GEOMETRIC DESIGN: PHASE I	6. Performing Organization Code
7. Author(s) Shaw-Pin Miaou, Patricia S. Hu, Tommy Wright, Stacy C. Davis, Ajay K. Rathi	8. Performing Organization Report No.
9. Performing organization Name and Address Center for Transportation Analysis Oak Ridge National Laboratory P.O. Box 2008 Oak Ridge, Tennessee 37831-6366	10. Work Unit No. (TRAIS) NCP Code 3A4a1102 11. Contract or Grant No. DTFH-61-90-Y-00036
12. Sponsoring Agency Name and Address Office of Safety and Traffic Operations, R&D Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296	13. Type of Report and Period Covered October 1990 - January 1993 Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

Contracting Officer's Technical Representative - Harry Lum (HSR-20)

16. Abstract

The purpose of this study was to establish empirical relationships between truck accidents and highway geometric design. First, statistical frameworks based on Poisson and negative binomial regression models were proposed. Preliminary models were then developed using accidents and road inventory data from the Highway Safety Information System (HSIS). Three roadway classes were considered in the model development: rural Interstate, urban Interstate and freeway, and rural two-lane undivided arterial. The maximum likelihood method was used for estimation of model parameters. Information criterion, asymptotic t-statistic, and goodness-of-fit test statistics were employed to evaluate the estimated models. The model results based on data from one of the HSIS States–Utah, were used for analysis and for suggesting areas in which the quality and quantity of the existing HSIS data can be enhanced to improve the developed models.

Despite the limitations in existing Utah data, some encouraging preliminary relationships were developed for horizontal curvature, length of curve, vertical grade, length of grade, shoulder width, number of lanes, and annual average daily traffic (AADT) per lane (a surrogate measure for vehicle flow density). Goodness-of-fit test statistics indicated that extra variations (or overdispersion) existed in the data over the developed Poisson models for all three roadway classes. Subsequent analyses suggested that a future study can be performed to enhance the predictive power of these preliminary models by including detailed truck exposure information, e.g., time-of-day, truck type, and weather conditions, by considering more explanatory variables, such as roadside design and superelevation, and by reducing the sampling errors of vehicle exposure data (both AADT and truck percentages).

17. Key Words	18. Distribution Statement
Highway Safety, truck accidents, highway geometric design, Poisson regression negative binomial regression, data needs	No restrictions. This document is available to the public through the National Technical Information Service Springfield, VA 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	152	

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	ml
gal	gallons	3.785	liters	l
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: Volumes greater than 1000 l shall be shown in m^3.				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5(F-32)/9 or (F-32)/1.8	Celcius temperature	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	l
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
psi	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	ac
ha	hectares	2.47	acres	mi^2
km^2	square kilometers	0.386	square miles	
		VOLUME		
ml	milliliters	0.034	fluid ounces	fl oz
l	liters	0.264	gallons	gal
m^3	cubic meters	35.71	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg	megagrams	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact)		
°C	Celcius temperature	1.8C + 32	Fahrenheit temperature	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	psi

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

(Revised August 1992)

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF TABLES (continued)

1. INTRODUCTION

The U.S. economy continues to depend heavily on commercial trucks for moving goods and materials. Trucking accounted for nearly one-third of all domestic intercity freight traffic and nearly three-fourths of the Nation's intercity freight bill in 1987.[1] The passage of the 1982 Surface Transportation Assistance Act (STAA) preempted more restrictive State vehicle size and weight limits, and has allowed longer and wider trucks to travel on a designated National highway network. Furthermore, States are expected to provide reasonable access beyond the national network to truck terminals and service facilities. Among several criteria, safety was identified by Congress as the primary criterion to guide the selection of the national network and the associated access roads.

Since the passage of the 1982 STAA, safety performance of large trucks has become a more important and pressing issue of public and government concerns. Safety questions, such as whether the current highway design is adequate to serve these larger trucks and what highway geometric conditions pose the most serious safety problems for them, are of primary interest. These questions can be better addressed if truck accident involvement rate, defined as the number of trucks involved in highway accidents per truck miles traveled, and accident probability can be accurately estimated for different truck configurations under different highway geometric conditions.

The need to establish statistical relationships between truck accidents and highway geometric design and the frustration among researchers to develop such relationships were properly described in a recent paper by Harwood, et al.:

> *The data . . . clearly illustrate the effect of two key variables related to hazardous materials routing--roadway type and area type--on truck accident rate. An attempt was made to determine the relationship between two traffic volume factors (AADT [annual average daily traffic] and percent trucks) and truck accident rate, but no consistent results were obtained. Consideration of the effects of additional geometric variables . . . on truck accident rates . . . would be desirable . . .. However, it should be recognized that the development of reliable relationships between geometric features and accidents is a difficult statistical task. Previous attempts . . . have had mixed results and no set of geometric-accident relationships is widely accepted.*[2]

Indeed, vehicle accidents are complex events involving the interactions of many factors, including not only the road, but also the vehicle, the drivers (human factors), the traffic, and the environment (e.g., weather and lighting conditions). Because some of these interacting factors are qualitative and stochastic in nature, e.g., drivers' behavior and weather conditions, the actual relationships between vehicle accidents and highway geometric design are inevitably empirical and statistical. To give a somewhat exaggerated example, a "perfectly" built highway can still be "unsafe" if it is highly congested or full of reckless drivers. Therefore, to establish such a relationship between truck accidents and highway geometric design, the analysis requires a statistical framework that is capable of modeling the inherent stochastic nature of the accident events, as well as characterizing the interacting effects of the associated factors. In addition, to obtain realistic results from the models, the analysis requires good accident, traffic, highway geometric, and environmental data, as well as good truck travel (or exposure) information, in both quality and quantity.

The purpose of this project was to develop preliminary statistical models for establishing the relationship between vehicle accidents involving large trucks and key highway geometric design variables by using the existing data sources.[1] The specific truck safety questions that the developed models are intended to address include:

- Given a section of highway, how safe is it for large trucks in terms of both accident rate and accident probability?
- Given a set of highway geometric design elements, which elements have relatively more impact on the safety performance of large trucks?
- What reduction in large-truck accident involvement rates can be expected from various improvements in highway geometric design?

Because of the possibility that existing data sources might not be able to fully support such a study, this project was to be conducted in two phases. In phase I, statistical frameworks suitable for describing such relationships were proposed. On the basis of the proposed statistical frameworks, preliminary models were subsequently developed using existing data sources. The primary data source used in this study was the Highway Safety Information System (HSIS), a highway safety data base developed by the University of North Carolina for the Federal Highway Administration (FHWA). The performance of these preliminary models was then statistically evaluated to determine whether the existing data were sufficient to develop such relationships with acceptable precision. In addition, a "data needs" study was conducted to identify additional data and explanatory variables for enhancing the predictive capability of the preliminary models. In the end, sample size requirements in terms of truck miles were estimated for developing representative "National" models to describe truck accident-geometric design relationships with some specified levels of statistical precision. Should the FHWA determine that the effort to

[1] In this study, a large truck is defined as a vehicle having a gross vehicle weight rating of 10,000 lb (4 535 kg) and over, where gross vehicle weight rating is the weight of a vehicle when loaded to its capacity.

collect these additional data and variables is cost effective, then phase II of the project would proceed with the objective of improving the preliminary models developed in phase I.

This report presents the research efforts and results from the first phase of this project. It includes an extensive literature review, the researchers' experience working with the HSIS data base, safety performance of large trucks based on the HSIS data, two proposed statistical frameworks for establishing such relationships, the numerical results and statistical implications of the preliminary truck accident models developed using the data from the HSIS, and a "data needs and sample size requirements" study.

Following this chapter, chapter 2 summarizes findings of the literature review. This chapter is intended to provide an overall perspective for the reader on the progress of truck safety research. It covers studies on the safety implications of various truck configurations and their relationships to various roadway, driver, vehicle, traffic, and environmental variables, as well as the existing policies, guidelines, standards, and practices relevant to highway geometric design. Chapter 3 describes the researchers' experience working with data of two selected States in the HSIS--Utah and Illinois. Major topics include the matching of truck accidents and road inventory data files; the availability of truck accident, highway geometric, traffic, and other relevant data; and summary statistics of the truck safety performance derived from the data by roadway class, truck configuration, and accident severity type. Chapter 4 proposes a Poisson regression based statistical model framework to establish relationships between truck accidents and key highway geometric design variables. The available highway geometric, traffic, and truck accident data from three roadway classes in Utah were used to develop preliminary models. Chi-square goodness-of-fit test statistics and an information criterion were employed to evaluate the significance and precision of the developed models. Chapter 5 presents an alternative statistical model framework -- a negative binomial regression based framework, to address the uncertainties associated with the models developed in chapter 4. Specifically, this model framework intends to quantify the effects of the following two major factors on the overall model uncertainty: (1) uncertainty on truck exposure data due primarily to sampling errors, and (2) "omitted variables" -- the explanatory variables for the occurrences of truck accidents that are not included in the models. Chapters 6 and 7 contain a "data needs and sample size requirements" study. These two chapters conclude the phase I study by suggesting areas and ways in which the quality and quantity of the existing data in the HSIS can be enhanced to improve the preliminary models developed in this study, and by estimating the sample size requirements for developing representative truck accident-geometric design relationships for the Nation.

2. LITERATURE REVIEW

INTRODUCTION

The main purpose of this chapter is to provide an overall perspective for the reader on the progress and the major findings of truck accident research. An extensive literature search and review was conducted to determine the magnitude of large truck accidents and the relationship to various roadway, driver, vehicle, traffic, and environmental variables (e.g., weather). Studies on the safety implications of various truck configurations were reviewed as well as the existing policies, guidelines, standards, and practices relevant to highway geometric design.

The literature search was conducted through a computerized search with the Transportation Research Information Services (TRIS) containing the Highway Safety Literature File and the Highway Research Information Abstracts. The following key words were used in this search: highway safety, truck accidents, highway geometric design, and accident methodology. The search covered all published work (included in the data bases) since 1970. The search identified several special reports from the Transportation Research Board (TRB), the Federal Highway Administration (FHWA), and the National Cooperative Highway Research Program (NCHRP). Using the references cited in these special reports and other publications identified in the TRIS search, several other relevant reports and publications were identified.

As would be expected, an abundance of literature was available since a great deal of research has been done in various facets of accidents involving large trucks. It included studies of truck accidents on different highway functional classes; accident rates of different severity types; relationships between vehicle configuration and accidents; truck accidents at interchanges, ramps, and work zones; and various other aspects of truck safety. The review of the available literature focused on obtaining information on the following key issues related to study objectives:

- What is the magnitude of large-truck accidents and how does it vary by highway functional class and by truck type?
- How is the analysis of truck safety issues typically performed?
- Can accident frequencies provide meaningful information? If exposure data are used, what are the measures of accident exposure for large trucks and how are the data obtained?
- What highway geometric design variables influence the safety record of large trucks and what relationships have been established to date?

- What is the relationship of truck operating characteristics (length, gross weight, etc.) to the truck accident experience?

In accordance with the key questions outlined above, the reviewed literature was organized into the following categories: (1) measuring truck safety: rates and frequencies, (2) magnitude of and trends in truck accidents, (3) truck configurations and accident rates, (4) accident experience by roadway type, (5) truck accidents and geometric design, and (6) driver and environmental factors in truck accidents. Significant research efforts and their findings are discussed in the following sections for each of these categories. A summary of the literature review is provided in the last section of this chapter.

MEASURING TRUCK SAFETY: RATES AND FREQUENCIES

Accident frequencies (counts) and rates are often analyzed to study the highway safety issues. The accident frequency data, which is readily available through statewide accident data bases or national data bases such as the Fatal Accident Reporting System (FARS) of the National Highway Traffic Safety Administration (NHTSA), are used in preliminary accident analysis or used in lieu of accident rates when accurate exposure data are not available. However, it is well known that the conclusions drawn from an analysis based on accident frequency alone could be misleading since the method does not take into account the likelihood of the event occurrence or the opportunities of accident involvement.(3)

On the other hand, the measures of exposure used in accident analysis are complex and have been the subject of much discussion and research. The available literature suggests a diversity of opinion about what is appropriate in a given situation. The need for better exposure data and appropriate exposure measures has often been identified.(4) Traditionally, however, exposure measures based on vehicle miles of travel and number of vehicles in operation have been used in accident research, simply because better, well-defined exposure measures do not exist. A comprehensive review of exposure measures used in accident analysis was provided by Council, et al. and Bowman and Hummer.(5,6)

The use of exposure measures in accident analysis poses conceptual and procedural problems when the objective is to analyze a specific accident type or accidents involving different vehicle types. The problems stem from determining an appropriate exposure measure that accurately reflects the likelihood of accident occurrence and one that can be derived from the commonly available data. The problem has been a topic of continuing research and has been addressed in different ways.

The traditional approach has been to categorize an accident by vehicle type and to estimate the exposure by determining the travel attributed to that type of vehicle. But as Khasnabis and Assar pointed out, in reality, the exposure to an accident for a particular vehicle type is caused not only by travel attributed to that type but also to the travel generated, in part, by all other types of vehicles in the traffic stream.(7) Therefore, the traditional approach attributes accidents involving more than one vehicle type to only one type of vehicle. For example, there were 70,000 accidents involving at least 1 truck in Michigan in 1982. These accidents involved approximately 76,000 trucks and 48,000 other vehicles (mostly passenger cars). Based on this data, an argument can be made that the exposure effect of non-truck vehicles should be taken into consideration while determining the truck accident rates.

Thorpe proposed a method of exposure measurement based on the idea of induced exposure.(8) The basic premise of this method is that the relative exposure for certain types of drivers, vehicles, driving environments, etc., can be determined from the not-at-fault representation of that analysis category in multivehicle accidents. The advantage of this approach is that exposure measures for a vehicle type, class of drivers, or driving environment can be determined from the accident data itself. That is, the induced exposure model permits the determination of exposure without obtaining vehicle counts and is, therefore, quick and economical. However, the method is based on the assumption of a large sample which limits its applicability to many situations, in particular location-specific problems. Also, Thorpe makes an assumption that single-vehicle accidents are caused entirely by the actions of the vehicles which are involved in the accidents. This assumption is subject to debate. Thorpe's approach was modified by Haight and Koornstra; both made more use of available detailed accident information.(9,10) Mengert assessed the validity of induced exposure models by applying the exposure estimates to actual accident data and suggested that the induced exposure methods did not provide credible estimates of exposure.(11)

Carr, Hall, and Cerrelli have considered models based on induced exposure that used the information on driver-at-fault as indicated by the officer investigating the accident.(12,13,14) Unlike strict induced exposure models, these models (often called quasi-induced exposure models) use information from the accident reports. Recent work by Maleck and Hummer suggests that quasi-induced methods can provide reasonably accurate exposure estimates and may be useful in many situations where exposure data are not readily available.(15)

Khasnabis and Assar discussed the complications in developing exposure measures when the objective is to separate accident data into two or more categories and suggested a technique that uses the concept of "interaction between vehicles."[7] The technique is based on categorization of accident data into accidents by vehicle types and the estimation of accident rates by taking into account the travel (vehicle miles of travel or VMT) contributed by not only the particular vehicle type but other interacting vehicles as well. An example use of the suggested technique was provided by Khasnabis and Assar through an analysis of heavy truck accident data for 1982 in Michigan.

Using an approach based on estimating the accident exposure for each accident type, researchers at the University of North Carolina (UNC) developed exposure measures for a variety of situations including accident studies for intersections, interchanges, roadway segments without intersections, and analyses dealing with fixed-object accidents and accidents involving specific vehicle types.[5] The first three situations involve "location-oriented" exposure measures, whereas the other two involve exposure measures in an entirely different context. These exposure measures were characterized by the authors to be "non-traditional" in that they deviated substantially from the standard vehicle miles of travel (VMT) concept. In general, they require more detailed traffic and highway geometric information, such as vehicle speed and vehicle flow by traffic lane, standard deviation of vehicle speed, and lane width. It seems that these measures can provide more insight for the microscopic research problems faced by highway safety researchers.

Although the "non-traditional" exposure measures, including the induced (or quasi-induced) exposure of Thorpe and the microscopic exposure of Council, et al., are interesting concepts, they have not yet been accepted as an effective measure for deriving vehicle accident rate. One reason is that they are harder to interpret and apply in practice than the standard VMT measure. Another possible reason is that vehicle travel by vehicle configuration and time of day can now be more accurately and cost-effectively obtained through the use of automatic traffic recording machines.

MAGNITUDE OF AND TRENDS IN TRUCK ACCIDENTS

Accident frequency and involvement rates of large trucks (gross vehicle weight rating (GVWR) of 10,000 lb (4 535 kg) and over) have been widely documented. Still, a complete picture of truck safety is not available due to lack of detailed data in many categories of interest (distribution of accidents by vehicle type and by weather condition, for example) and lack of exposure data in general. Also, statistics are often developed using statewide or regional data bases which cannot be used to estimate the national trends. In this section the accident experience of large trucks is described first by the frequency of truck involvement in accidents and then by observed accident rates. The involvement of trucks in accidents relative to passenger cars is discussed as appropriate and the trends in accident involvements or rates are indicated where possible.

Frequency of Truck Accidents

Trucks (single-unit and combination) were found to be involved in approximately 6 percent of all reported accidents but account for 12 percent of all fatal accidents.[16] A recent study by Campbell et al. indicated that of the 11,069,000 vehicles estimated to have been involved in accidents in the United States in 1986, 2.1 percent (235,000) were combination trucks, accounting for 6.7 percent of fatal accidents.[17] In comparison, nearly 75 percent of all accidents and 60 percent of fatal accidents involve passenger cars. Each year approximately 4,800 persons are killed in accidents involving trucks, and almost 75 percent of these fatalities are occupants of non-truck vehicles.[18] The higher proportion of fatal accidents involving trucks can be explained, in part, by the difference in size and weight between trucks and non-truck vehicles.

Based on FARS data, both the number of fatal accidents and the proportion of all fatal involvement for trucks increased slightly between 1975 and 1985.[18] Similar trends have been observed for truck involvement in all accidents by NHTSA's National Accident Sampling System (NASS), but the estimates are not precise due to small sample size for trucks.[19] Analysis of data from Michigan also suggests that accidents involving trucks are increasing annually after a low point reached in 1982.[17,20]

Truck Accident Rates

Accident rates involving trucks have traditionally been calculated in two ways: (1) truck accident involvement rate: number of trucks involved in accidents divided by total truck miles, and (2) truck accident rate: number of accidents involving trucks divided by total truck miles. Depending on the availability of accident and truck exposure data, these rates have been categorized by truck configuration, roadway type (or functional class), and accident severity type.

The total truck accident rates have been estimated to be anywhere from 0.7 to 7.6 per million truck miles (MTM) depending upon the type of roadway (number of lanes, divided vs. undivided, Interstate vs. noninterstate), location (urban vs. rural), and vehicle configuration.[21, 22, 23, 24, 25] The nationwide average truck accident rate is estimated at 3.3 per MTM for combination trucks.[19] Fatal accident involvement rate for trucks, for which generally better data are available through FARS and other national data bases, is estimated to be 5.5 per 100 MTM for combination trucks and 3.3 per 100 million vehicle miles of travel (MVMT) for all vehicles.

Although trucks are found to have slightly less than half the accident involvement rates of passenger cars (2.87 per MTM vs. 6.33 per MVMT), they were found to be overrepresented in fatal accident involvements.[1] Meyers examined accidents on 34 limited access toll facilities for the 1976 through 1978 period, using the detailed exposure data available from toll receipts.[26] The analysis of 73,500 accidents with almost 50 billion vehicle miles of travel indicated that trucks with GVWR between 10,000 and 26,000 lb (4 535 and 11 791 kg) were involved in 2.35 times more fatal accidents than passenger cars. The fatal accident involvement rate for heavy trucks (over 26,000 lb (11 791 kg)) was almost twice the rate for passenger cars. Jovanis and Delleur, as well as Preusser and Stein, analyzed truck accidents on toll facilities to take advantage of the exposure data available from the toll receipts.[24,27] The analysis of 3 years of data from four facilities in the latter study indicated that tractor semitrailer trucks were overrepresented in multivehicle accidents, fatal accidents, and total accidents per mile of travel when compared with passenger cars.

Several other researchers have used large national or statewide data bases, e.g., Trucks Involved in Fatal Accidents (TIFA) developed by the University of Michigan, to calculate truck accident rates. The general finding of these studies is that the accident involvement rates for trucks are lower than passenger car rates for all accidents but considerably higher for fatal accidents. The higher fatal accident involvement of trucks can be explained, in part, by the higher share of truck travel on rural roads where speeds are high and accidents tend to be more severe. On the other hand, the lower overall involvement of trucks in accidents can also be attributed to the fact that truck travel generally occurs on low risk roads.

Truck accident involvement rates are increasing annually after a low reached during the recession in early 1980's.[20,28,29] According to NHTSA's involvement data for trucks and FHWA travel estimates by functional class, this pattern holds for both Interstate and noninterstate roads.

TRUCK CONFIGURATIONS AND ACCIDENT RATES

Several studies have compared accident rates for different truck types with most attention given to the comparison of accident experiences of tractor semitrailers and twin trailer trucks. An extensive review of studies comparing relative accident records of twin trailer trucks and tractor semitrailers was conducted by the TRB for its special report 211---Twin Trailer Trucks.[25] TRB Special Report 223 presents yet another comprehensive review of studies on this subject matter while focusing mainly on research that was performed after the TRB Special Report 211.[1] The key findings of some of the major studies are highlighted below.

A 1988 study by the University of Michigan Transportation Research Institute (UMTRI) found that twins have a 10 percent higher fatal accident rate nationwide than tractor semitrailers after adjusting for differences in travel by road class, time of day, and area of truck travel.[17] Using a case-control methodology, a 2-year study by Stein and Jones found that compared to semitrailers, twins are involved in two to three times more accidents regardless of accident type, truck operating characteristics, driver characteristics, and environmental and road conditions.[30] However, Stein and Jones' results were found to be questionable in that they might have undercounted the amount of twins relative to tractor semitrailers in their exposure population, thereby inflating the involvement rates of twins relative to tractor semitrailers.[31] In contrast, a recent study by Jovanis, et al. found that twins had lower accident involvement rates than semitrailers on Interstate, State, and local roads.[32] The study was based on an analysis of 3 years of accident and exposure data. Graf and Archuleta found that twins have higher accident rates than semitrailers on rural roads, but a lower involvement on urban roads.[33] When the environment is not taken into account, no statistically significant differences in accident rates between the two vehicle configurations were observed. Similar results were found in earlier studies by Glennon, Chira-Chavala, et al., and Yoo, et al.[34,35,36] Based on a synthesis of prior studies, TRB Special Report 211 concluded that twins are more likely to be involved in accidents than semitrailers, but indicated that on a ton-mile basis there are no significant differences between the two vehicle classes.[25]

Very few studies have compared the accident involvement of straight trucks with other vehicle configurations (twins or tractor semitrailers). A study of the Bureau of Motor Carrier's Safety (BMCS) accident data bases determined that most types of doubles and singles have higher accident involvement rates than the straight trucks, although loaded straight trucks were more involved than empty ones.[37] The data also suggest the straight trucks are involved in one out of every four truck accidents.[38] However, no information was provided on the relative exposure of the two types and thus no comparison of accident rates. The relative fatal accident involvement rates, based on FARS, for three truck types in year 1985 were also summarized in TRB Special Report 225: (1) single unit: 7.7 involvements per 100 MTM, (2) tractor semitrailers: 10.2 involvements per 100 MTM, and (3) twins: 11.2 involvements per 100 MTM.[31] Also, twins were found to have been involved in proportionally more fatal rollover and jackknife accidents than tractor semitrailers, but proportionally fewer fatal multiple-vehicle accidents.

In summary, the research work to date suggests minor differences in overall accident rates between twin-trailer trucks and tractor semitrailers, the two truck types which are often compared. However when the accident severity and type of roadway are taken into consideration, some of the differences in accident rates of different vehicle configurations are significant. It should also be noted that calculation of exposure data is problematic in most of the studies and therefore a definitive assessment of the comparative accident rates is not possible. Even when an attempt is made to account for exposure of different vehicle configurations through controlled experiments, the study designs are such that the results cannot be extrapolated to a broad range of situations.

ACCIDENT EXPERIENCE BY ROADWAY TYPE

Do the accident and severity rates vary by the highway functional class? What roads have the worst safety records for trucks? Is there a significant difference between truck accident experience on rural vs. urban settings? The available literature was reviewed to answer these questions and also to see how the accident data should be categorized by road type in this study.

Previous studies on large-truck accidents generally agree on the types of road that have the highest likelihood of accident involvement. An UMTRI study found that the highest fatal accident rates for large trucks were on undivided rural primary highways and the lowest rates were on divided highways.(17) The higher accident rates on rural highways were attributed to relatively high speeds of travel and inadequate geometric design standards on these roads. A 1977 study by Hedlund found that truck accidents on two-lane rural roads are more likely to be fatal than those on 4-lane rural roads.(38) Graf and Archuleta found that fatal accident involvement for large trucks is higher in rural areas, but total accident involvement is higher in urban areas.(33)

Based on the research findings to date, it is quite clear that the roadway type plays an important role in the frequency and severity of truck accidents. The studies generally agree on two findings: (1) divided roads are safer than undivided roads for truck travel; and (2) rural noninterstate roads account for the majority of large-truck fatal accident involvements, and have the highest involvement rate per mile traveled. Both of these relationships are consequences of relative geometric design standards of the highway sections. The differences in accident experiences on urban vs. rural settings are also identified but the relationships are not yet well established.

In developing truck accident rate models for hazardous materials routing, Harwood, et al. derived truck accident involvement rates for 9 roadway types in 3 States. Table 1 shows their results.(2) Two observations can be made from the table: (1) urban highways have consistently higher truck accident involvement rate than their rural counterparts, and (2) urban multilane arterial and one-way streets have the highest truck accident involvement rates among all roadway types.

Table 1. Truck accident involvement rates
(number of trucks/million truck miles) by roadway type and State.[2]

Roadway Type	California	Illinois	Michigan
Rural Two-Lane Arterial	1.73	3.13	2.22
Rural Multilane Undivided Arterial	5.44	2.13	9.50
Rural Multilane Divided Arterial	1.23	4.80	5.66
Rural Interstate	0.53	0.46	1.18
Urban Two-Lane Arterial	4.23	11.10	10.93
Urban Multilane Undivided Arterial	13.02	17.05	10.37
Urban Multilane Divided Arterial	3.50	14.80	10.60
Urban One-Way Street	6.60	26.36	8.08
Urban Interstate & Freeway	1.59	5.82	2.80

TRUCK ACCIDENTS AND GEOMETRIC DESIGN

Geometric design elements (lane width, horizontal curvature, vertical grade, etc.) of the roadway sections play a pivotal role in safe operation of motor vehicles. The highway facilities are designed to accommodate the largest design vehicle likely to use that facility with considerable frequency. Because passenger cars account for over 85 percent of travel on the Nation's highways, the roadway system has many geometric design standards that are inadequate for large trucks. With the increase in allowable maximum dimensions for truck tractor trailers permitted by the 1982 Surface Transportation Assistance Act (STAA), even portions of the Interstate highway system have become deficient in geometric design. The existence of restrictive geometry limits the operation of the 1982 STAA vehicles and can have a detrimental impact on their safety record.

The relationship between highway geometric design features and vehicle accidents has been addressed in numerous past studies. The NCHRP Report 197 contains an excellent summary of research performed up to 1978 including a synopsis of major findings from over 400 reports and publications.[39] Unfortunately, most of this research does not distinguish between trucks and non-trucks. The findings of the literature review in NCHRP Report 197 suggest that the following geometric design variables of highways have significant influence on the accident experience:

- Number of lanes and lane width.
- Shoulder surface type and width.
- Median width and type of median barrier.
- Horizontal and vertical alignment.
- Intersections.
- Roadside features.

As indicated above, most of these findings were not directly applicable to accidents involving trucks. Also, a clear cut relationship was not established in all cases. For example, a 1973 study by Garner and Deen concluded that as the median width increases on freeways, the accident and severity rates decrease.(40) However, an earlier California study by Moskowitz and Schaefer found that median width is unrelated to the total accident and injury accident rates.(41) Note also that the study findings vary between highway functional classes.

The TRB Special Report 223 contains an overview of the safety record of large trucks and provides a review of current literature on various aspects of accident experience of combination trucks: involvement in accidents and accident rates, vehicle configuration and accident rates, vehicle configuration and severity of accidents, and accident experience by road type and characteristics.(1) Many of the studies cited in this report are also examined in the TRB Special Report 223. On the basis of an earlier TRB study, the TRB Special Report 223 suggests that the following six highway design features may have significant bearing on the accident experience of large trucks (termed STAA vehicles for vehicles permitted under the 1982 STAA): lane width, shoulder width and type, bridge width, roadside and sideslopes, pavement edge drops, horizontal curves, and intersections.(1)

Three major conclusions can be drawn from the review of literature cited in these reports. First, very few research efforts have concentrated on the relationship of truck accident rates to highway geometric design variables. Second, most of the work to date has emphasized one particular geometric design variable (e.g., lane width) rather than simultaneously analyzing all variables of design which may have significant influence on the accident experience. Finally, exposure measures and calculations thereof are often questionable in these studies.

The only identified study which related truck accident rates to several highway geometric design elements in a manner quite similar to that desired by this study was performed by Joshua and Garber.(42,43) The overall objective of that study was to identify those traffic and geometric variables that are significantly associated with accidents involving large trucks. The study used

1984 to 1986 data on truck accidents for 43 selected sites in Virginia which had a high number of truck accidents. The exposure data were derived through vehicle counts at the selected sites and the geometric data were obtained through on-site measurement of lane width, horizontal alignment, and vertical alignment. Statistical relationships for three highway environments were then developed for estimating the number of truck accidents using a set of traffic and geometric variables. According to Joshua and Garber's models, the significant geometric design variables were slope change rate (SCR) for primary highways and curvature change rates for freeway type facilities.

The major limitations of Joshua and Garber's study lie in the relatively small (and perhaps biased) sample size, the lack of consideration of driver or environmental factors, and the lack of distinction between various truck types. While one can argue the limitations of the analysis methodology used (see chapter 4 for detail), the study is indeed an early attempt towards quantifying the relationship between truck accident rates and highway geometric design variables. The measures for the complexity of highway alignment (horizontal curvature and vertical grade) proposed by Joshua and Garber also represent a significant contribution towards the quantification of geometric complexities in nonhomogeneous highway sections.

The remainder of this section summarizes the current knowledge about the relationship between vehicle accidents and these geometric design elements, with emphasis on truck accidents. Most of these findings are based on the research cited in NCHRP Report 197 and TRB Special Report 223.[39,1] The intuitive relationship between a particular geometric design element and truck operating characteristics is discussed as applicable.

Number of Lanes, Lane Width, and Surface Type

The number of lanes may not have significant influence on accident rates on rural highways but may be important in urban environments. An increase in number of lanes on freeways and at intersection approaches has been found to be associated with reduced accident rates. This is, however, assuming that vehicle volume does not increase after the change and as a result vehicle density is reduced.

The impact of lane width on accident rates has been well researched and documented. Adequate lane width is important to provide sufficient lateral clearance between vehicles moving in the same or opposite directions. Research findings suggest that accident rates decrease as lane widths increase for rural highways.

Lane width has been identified as a principal factor affecting bridge safety. Wider lane and pavement widths on bridges result in significantly lower accident rates. Narrow bridges and bridge approaches on downgrades or sharp curves, can increase the risk of accidents and may pose a greater problem for large trucks than passenger cars.

Shoulder Surface Type and Width

Adequate shoulder width is important for increasing the chances for safe recovery when vehicles run off the road. Shoulder width is particularly important on sections of highway with sharp horizontal curves. Since trucks are overrepresented in single-vehicle, run-off the road accidents, adequate shoulder width is critical to the safety and operation of large trucks.

Previous research suggests that as the shoulder width on rural two-lane highways increases, the accident rate decreases. However, for freeways, the right shoulder width was found to be unrelated to accident rates. It was suggested that the shoulder widening, therefore, is not as effective as lane widening.

Shoulder type (paved vs. unpaved) is also important in terms of stability of vehicle when it goes off the road, particularly for large trucks. Roadways with paved shoulders are found to have lower accident and severity rates than similar highways having unpaved shoulders of the same width.

Median Width and Type of Median Barrier

The increase in median width may reduce accident and severity rates by allowing the vehicles more recovery time and reducing severe accidents where a vehicle crosses the median and collides with the vehicles from the other direction. However, the difference may be insignificant. Previous research is inconclusive on the effect of median width on accident rates. Some studies suggest a significant decrease in accident rates when median width is increased. However, other findings suggest that width of the median is unrelated to the accident experience.

Insofar as the type of median is concerned, nontraversable median barriers are found to decrease the fatal accident rates but increase all other types of accident rates, including the overall accident rates. The cable type median barriers have higher total and fatal accident rates than solid beam or concrete type barriers. While it is generally agreed that the median barriers are not designed to handle large trucks, the impact of median type on truck accident involvement has not been studied.

Horizontal Curves

Accidents are more likely to happen on horizontal curves than on straight sections because of the vehicle dynamics and additional effort required by the driver to control the stability of the vehicle. As the degree of curvature increases, the accident rate increases. Previous research indicates a nearly linear relationship between the accident rates and the degree of curvature. The effect of the length of curves on truck accidents has not yet been well quantified.

Trucks are more sensitive to horizontal curves than passenger cars because of their higher center of gravity and have particular difficulty negotiating sharp curves because of the off-tracking problems. Sharp curves are often found in areas with topographical constraints. Often, these curves are accompanied by narrow lane and shoulder widths and steep grades. A combination of these three design features poses a significant risk to the truck operation and safety.

Vertical Grades

Intuitively, there is a greater risk of accidents on grades than on straight sections. For trucks, downgrades may be more dangerous than upgrades because of their inferior deceleration capabilities compared to passenger cars. Upgrades, on the other hand, pose a different type of safety problem. Since trucks cannot maintain normal, prevailing traffic speeds on upgrades, the drivers of passenger cars occasionally attempt to perform unsafe passing maneuvers on two-lane roads. Also, substantial speed reductions on upgrades may lead to sudden braking by the following vehicles resulting in overturning or rear-end accidents. Research findings suggest that the accident and severity rates increase on grades. The effect of the length of vertical grades on truck accidents has not yet been adequately quantified.

Intersections

A significant portion of accidents occur at intersections--nearly half of the accidents on urban highways and one third of the accidents on rural highways. Large trucks have difficulty turning at intersections in many instances because of off-tracking and sight distance problems. The relationship between intersection geometrics and truck accidents has not been adequately quantified because so many other factors (e.g., traffic control) affect the accidents at intersections.

Roadside Features

Numerous design features are associated with the roadside: slopes, ditches, pavement edge drops, guardrail, frontage roads, access control, and clear roadside area. Since large trucks are more likely to be involved in single vehicle accidents than are passenger cars, adequate recovery areas at the road edge that are smooth, without steep sideslopes, and clear of fixed objects are important to truck safety. No clear cut relationships have been established relating roadside features to truck accident rates. This may be, in part, due to inadequate information in accident data bases or roadlog files.

The relationship between the geometric design features of highways and their safety records has been the subject of much research in the past. The research findings to date suggest that with a few exceptions, it is difficult to establish clear cut relationships between restrictive highway geometry and accident experience. Furthermore, the relationship between restrictive geometry and truck accidents has rarely been explored. Based on this review and given the particular handling and performance characteristics of large trucks, the key geometric features that may affect the safety records of these vehicles are identified as follows:

- Lane width.
- Shoulder width and type.
- Horizontal and vertical alignment.

DRIVER AND ENVIRONMENTAL FACTORS IN TRUCK ACCIDENTS

An accident is a random event, often the result of a complex interaction of several factors: the road, the drivers, the vehicle, the traffic, and the environment. The discussion so far has emphasized the vehicle and roadway factors contributing to the truck accidents. However, the driver and environmental factors also have significant influence on truck involvement in accidents. Driver factors in truck accidents include age, education and training, experience, fatigue (hours of driving), and other conditions such as alcohol or drug use. Environmental factors include pavement conditions, weather, lighting (time of day), and other traffic conditions.

The effect of driver and environmental factors for traffic accidents has been examined in a number of past studies. Specific to truck accidents, Eicher, et al. indicated that young drivers are involved in a disproportionately high share of large truck accidents.[16] Only 15 percent of large truck drivers are under the age of 30 but they account for nearly 30 percent of large truck involvement in accidents. Stein and Jones' analysis of relative risk of accident involvement for various drivers characteristics indicated three dominant driver factors in a truck accident: driving time, driver's age, and the carrier operation (e.g. long haul and short haul).[30] Drivers under 30 years of age were found to have significantly increased relative risk (1.8 to 1.0) of accident involvement over drivers over 30 years of age.

Time and environmental factors in truck related accidents have also been analyzed by many researchers. For example, a 1978 analysis by Cassidy identified time and environmental factors in truck-involved fatal accidents on a nationwide basis.[44] This study of 2 years of FARS data determined that (1) the highest percent of fatal accidents involving trucks occurs in July and the lowest occurs in February, (2) Friday is the peak day of the week for truck-involved fatal accidents, and (3) the proportion of heavy truck involvement in fatal accidents increases to 10 percent in rainy weather and to 16 percent in sleet or snowy weather compared to 8 percent involvement in all accidents. Some of these findings parallel the travel pattern and thus can be explained by relative exposures of the vehicles involved in the accidents. However, many other temporal trends were determined to be different than those exhibited by the FARS data. A study by Vallette, et al. suggested different trends for accidents occurring on different weekdays than those exhibited by the BMCS or FARS data.[45] In that study, accident and exposure data were collected over a 1.5-year period during 1976 to 1978 for 1,058 mi (1 703 km) of roadway in six States. The Vallette study suggested a uniform distribution of accidents on weekdays.

Jovanis and Delleur's study of truck accidents on toll roads in Indiana found that (1) truck accident rates were similar for day and night conditions, (2) snow was a significant weather factor increasing accident rates for all vehicle types, and (3) large truck involvement rates were lower during rainy weather conditions. However, an analysis of 1980 BMCS data by Chira-Chavala, et al. indicated that weather and light conditions have a small effect on fatality and injury accident rates for different truck types.[35] The relative frequencies of truck-involved fatal accidents were also discussed in the UMTRI presentation.[37] It was concluded that: "dawn stands out as having a distinct mix of accidents. Many of these seem to be related to fatigue; but the fatigue cannot be attributed to long hours of driving."

In summary, while it is widely recognized that the driver and environmental factors also have significant influence on truck involvement in accidents, the relationships have not yet been well established. Also, except in few instances, the studies have indicated conflicting trends in accident patterns in relation to driver and environmental factors. For example, the differences in daytime vs. nighttime truck accident rates have not yet been fully analyzed. Furthermore, many of the conclusions were based on studies with problematic data or analysis technique.

SUMMARY AND FINDINGS

The involvement of trucks in highway accidents has been a subject of much research and analysis. The emphasis of these studies has been on the comparison of safety records of trucks with passenger cars or relative safety records of various truck configurations for different accident types and severity, accidents on different road types and/or operating environments. The results of relevant previous and ongoing research were reviewed to determine the magnitude of large-truck accidents and the relationship to various roadway, driver, vehicle, traffic, and environmental variables (e.g. weather) with the emphasis on identifying the variables of interest for the analyses planned in this study.

Three conclusions were apparent from this review. First, although much research has been performed on various facets of accidents involving trucks, very little emphasis has been given to establish relationships between truck accidents and geometric design elements of highways. Not surprisingly, this review identified only one study which focused exclusively on the relationship between the truck accidents and highway geometric design elements. Second, a variety of analysis methodologies have been used to quantify the truck involvement in accidents. This is mainly because of different study objectives but also because of lack of proper exposure data. Data were often limited in many of the studies, making their conclusions limited (biased in some instances) in scope. In particular, the calculation of exposure measures is subject to debate. Third, the accident involvement rates and relationships derived in these studies are variable and often contradictory.

It should, however, be noted that despite the analytical and data limitations, significant progress has been made towards understanding the safety implications of trucks on the highways. The relative accident involvement of twins and semitrailers has been extensively analyzed. While the relationship between various factors contributing to truck involvement in accidents has not yet been well established, a number of efforts are currently underway to improve the data quality and analysis capabilities. With these caveats in mind, the key findings from the review of literature presented above are as follows:

- Exposure data for truck travel, central to calculating accident rates, are limited in both the quality and quantity. Therefore, a clear picture of the magnitude of truck accidents and a breakdown by several categories of interest (vehicle types, environmental factors, driver factors, etc.) are not available.

- Trucks are involved in a relatively small share of all accidents (nearly 6 percent) but in a relatively higher share (nearly 12 percent) of all fatal accidents. Also, large trucks have lower accident involvement rates (2.87 per MTM) than passenger cars (6.33 per MVMT) for all accidents but considerably higher rates for fatal accidents.

- Given the operating characteristics of large trucks and accident experience reported to date, lane width, shoulder width and type, horizonal curvature, steepness of vertical grades have been determined as key highway geometric design elements which may have significant influence on truck involvement in accidents.

- Accident rates for trucks are higher on undivided highways than on divided highways. Rural noninterstate undivided highways have the highest total accident and fatal accident involvement rates for trucks.

- Truck drivers under the age of 30 are involved in a disproportionately high percentage of accidents. Driving time and carrier operation are also significant factors in determining the relative risk of truck accidents.

- Several other factors influence truck accident and severity rates. This includes weather condition, pavement condition, location, time-of-day, etc. However, the relationships have not yet been well established.

3. DATA AND SUMMARY STATISTICS

INTRODUCTION

In this study, data from the Highway Safety Information System (HSIS), an accident data base developed by the Highway Safety Research Center (HSRC) of the University of North Carolina (UNC) for the Federal Highway Administration (FHWA), are employed for developing relationships between truck accidents and key highway geometric design variables. Because truck exposure (or truck travel) data in HSIS are currently not broken down by truck type, another data source, Highway Performance Monitoring System (HPMS), is employed as a supplementary data source whenever exposure data by truck type are needed.(46) HPMS has been implemented by the FHWA to assess highway systems by continually monitoring their physical condition and use. More specifically, HPMS is a data collection effort designed to provide current statistics on the mileage and use of highways, and to evaluate highway programs by monitoring changes in highway characteristics and performance.

The main purposes of this chapter are to give an overview of the HSIS and to describe the researchers' hands-on experience working with data from two selected States in the HSIS--Utah and Illinois. This includes discussions on data and file management, data and variable availability, and truck safety implications from the HSIS data. The focus of the discussion will be on Utah's data because they were used to develop models in the following two chapters. The discussion on Illinois data will be brief, since Illinois data were dropped from the final analysis in the report. The Illinois data were not included because it was discovered in the early stage of the model development that there were several undesirable data limitations.(47)

The chapter is organized as follows. The next section gives an overview of the HSIS. The *File Merging and Data Editing* section describes file merging procedures for linking truck accident files with the road inventory data. A summary of the highway geometric design and traffic variables that are currently available for Utah and Illinois in the HSIS and their associated summary statistics are presented in two sections entitled: *Available Variables* and *Summary Statistics of Utah Data*. The next section, *Truck Accident Frequency and Involvement Rate in Utah*, presents truck accident frequency and accident involvement rates in Utah for three roadway classes, two truck configurations, and three accident severity types. A summary on the HSIS data is presented in the last section.

HSIS OVERVIEW

HSIS is a prototype data base integrating all the data into one single system to facilitate highway safety research. Data files on vehicle accidents, highway geometric design, traffic conditions, and other relevant information collected by States were acquired, reviewed, quality checked, and rearranged in a unified SAS™ format for each individual State. Because different variables are collected and, occasionally, different definitions are used for variables with the same names collected by individual States, the data in HSIS for different States are not ready for direct merging.

At present, HSIS contains information on 5 States: Illinois, Maine, Michigan, Minnesota, and Utah. These States were the first to be selected for inclusion in HSIS because they were believed to have better accident and roadway data in terms of quantity and quality, as well as the ease of linking accident and road inventory data. Of these 5 States in HSIS, Utah and Illinois were considered to have the most complete information on highway geometric design, especially on horizontal curvature and vertical grade.[48] A detailed description of the Utah and Illinois data is documented in Council and Hamilton, and Council and Williams.[49,50]

Currently, the Utah data are stored in six separate files: roadlog, horizontal curvature, vertical grade, accident, vehicle, and occupant files; and the Illinois data are maintained in four files: roadlog (including horizontal curvature and vertical grade), accident, vehicle, and occupant files. These files have to be linked before any analysis can be performed. Figure 1 gives some example data items that are currently collected in each Utah file. Both road and accident related files are maintained on an annual basis so that year-to-year changes on highway geometric design and traffic conditions are recorded and vehicle accidents in a given year can be matched to the road inventory file of the same period. For the current analyses, accident and road related files for both States were acquired for the following periods:

- Utah: Accident related files were available annually from 1985 to 1989; roadlog file was available on an annual basis from 1985 to 1989; horizontal curvature file was only available for year 1987; and vertical grade file was only available for two years -- 1987 and 1990.

- Illinois: Accident files were available from 1985 to 1987, and roadlog files were available for the year 1987.

Since horizontal curvature and vertical grade usually change very little over the years, in the Utah roadlog file the 1987 horizontal curvature data were used for all 5 years of road sections, the 1987 vertical grade data were used for the first 3 years (1985-1987) of road sections, and the 1990 vertical grade data were used for the 1988 and 1989 road sections.

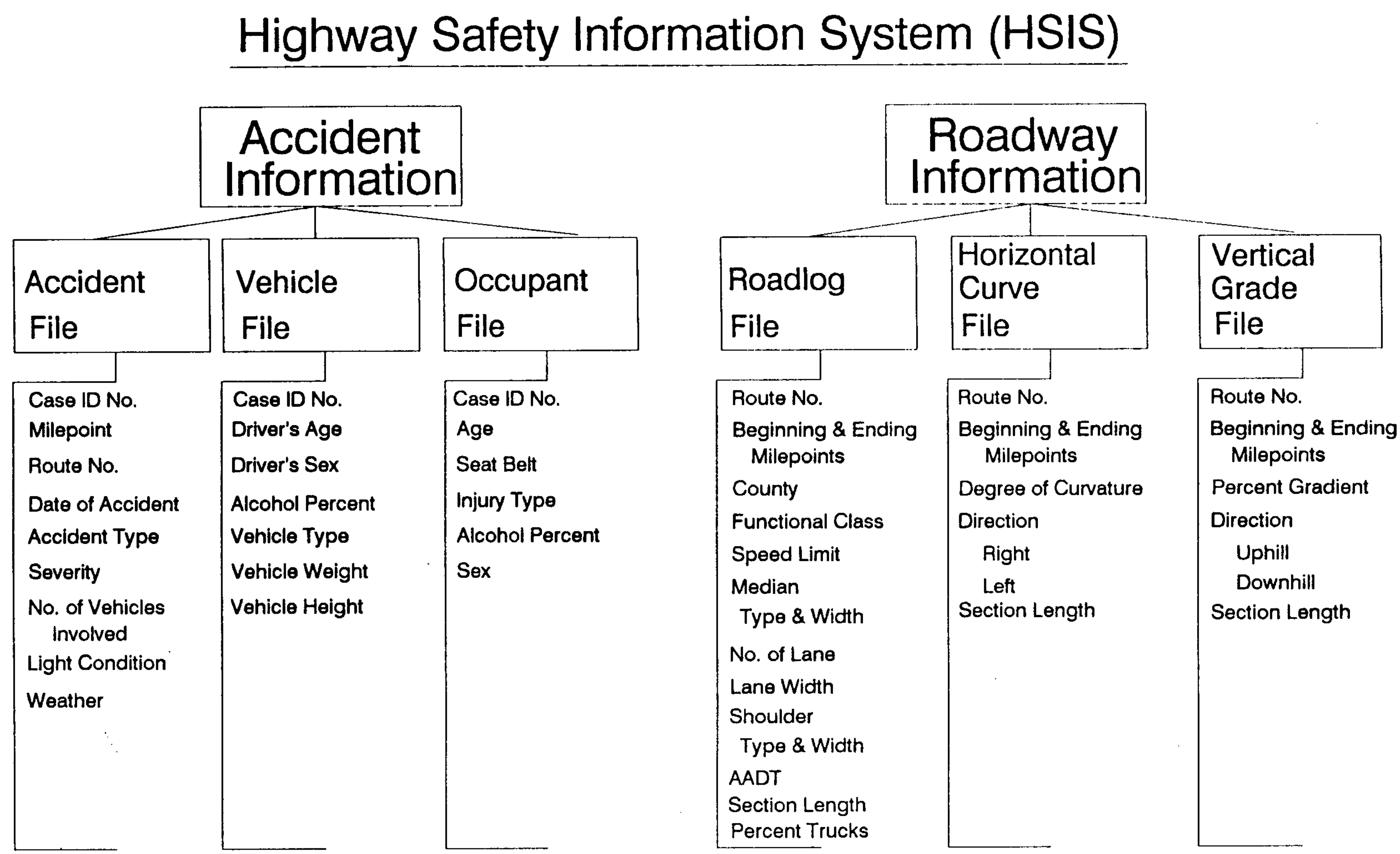

Figure 1. File and structure and example data items.

(Note that, as of September 1992, the HSIS is updated to include accident and roadlog data from 1985 to 1990 for all five States.)

FILE MERGING AND DATA EDITING

The key variable used in linking three accident files, i.e., accident, vehicle, and occupant files, is a unique identification number for each accident case called YRCASE and CASENUM, respectively, for Utah and Illinois. The key variables used to link the three road files (i.e., roadlog, curvature, and grade files) in Utah and between the accident file and the roadlog file in both states were the route number and milepoint where accidents occurred, and the route numbers and milepoints where a road section, a curve, and a grade began and ended.

Since the objective of this study is to determine the accident involvement rate and accident probability of a road section, given its geometric design, traffic, and other relevant characteristics, a road section (not the accident) is considered as a unit in the model development. Each accident record, which contains information from accident, vehicle, and occupant files such as accident type, accident severity, vehicle type, time of the accident, and driver(s)' conditions, is "mapped" onto the road section in the roadlog file where the accident occurred.

Each record in the roadlog file represents a homogeneous road section in terms of its cross-sectional characteristics, such as number of lanes, lane width, shoulder width, median type and width, and annual average daily traffic (AADT). Thus, for example, once the lane width changes on a particular road section, this section, together with its neighboring sections, is redelineated to reflect the change.

In the remainder of the section, file merging and data editing issues pertinent to Utah and Illinois are discussed separately as follows.

Utah

In the Utah roadlog files, each road section is not necessarily homogeneous in terms of its horizontal curvature and vertical grade. On the other hand, each road section in the Utah horizontal curvature and vertical grade files was homogeneous in terms of its horizontal curvature and vertical grade, respectively, but not in terms of other road characteristics. Therefore, when the road sections in horizontal curvature and vertical grade are "mapped" onto road sections in the roadlog file, each road section in the roadlog file may contain multiple (i.e., more than one) curvatures or grades.

In the vehicle file, vehicles involved in accidents are categorized into 37 vehicle types (or body styles). The truck category of interest includes: single-unit truck (excluding pickups and vans), truck trailer, tractor semitrailer, truck with two short-trailers, truck with one long-trailer, and five additional categories on tractors with two or more trailers. These truck categories typically have a gross vehicle weight rating greater than 10,000 lb (4 535 kg).[1] "Bobtails," i.e., tractors without trailers, are recorded separated from the other trucks. Note that "bobtails" are not included in this study because these accidents may require some special analysis.

Three relevant issues arose when preparing Utah data for analysis:

- Of the approximately 60,000 mi (96 600 km) of highway in Utah, the roadlog file contains road inventory of over 50,000 mi (80 500 km) of road system in which about 37,500 mi (60 375 km) are categorized as "local" routes. These local routes are categorized in "zones" which lump together all the mileages on different local roads within a given area. The beginning and ending mileposts of these local routes have no meaning in terms of road section length. It was suggested by Mr. F. Council that these local routes not be used for any analytical purpose. These local routes were therefore dropped before file merging was begun in this study.

- As indicated earlier, each road section in the Utah road inventory file was relatively homogeneous in terms of general cross-sectional characteristics and traffic conditions, but not necessarily in horizontal curvature and vertical grade. Therefore, each road section in the road inventory file may contain multiple horizontal curvatures or vertical grades. Two ways of resolving this problem were considered. One way is to create surrogate measures to characterize curvature and grade conditions along the length of each road section in the roadlog file. For example, Joshua and Garber have used the absolute changes or section-length weighted averages of curvatures and grades along the length of each section in the roadlog file.(42) Detailed discussion of these measures is included later in this section. Another way is to disaggregate those road sections with multiple curvatures and grades into smaller subsections in such a way that each subsection contains a unique set of horizontal curvature and vertical grade. While the former is considered less direct from the engineering point of view and may be difficult for design engineers to incorporate these measures into their current practice, the second method is considerably easier to interpret in a design context. Both approaches were considered in this study.

- There were five duplicate accident records due to coding errors. These coding errors were corrected before model development.

Some data editing was found to be required when using Utah roadlog data. The following editing was performed: (1) missing truck percentage and AADT were estimated based on the data from a neighboring section along the same route; (2) the whole route was deleted if none of the road sections on that particular route contain AADT or truck percentage information; (3) sections with an average truck volume of less than 200 trucks per day on Interstate and urban highways and 100 trucks per day on rural arterial were removed from analysis; and (4) routes with

[1] Gross vehicle weight rating (GVWR) is defined as the weight of a vehicle when loaded to its capacity.

no lane width information were deleted. Only a small fraction of the road sections was deleted as a result of this editing.

Horizontal curvature and vertical grade

Horizontal curvatures and vertical grades were coded in degrees per 100-ft (30.5-m) arc and percent, respectively. In addition, positive values indicated "right turn" and "upgrade" whereas negative values indicated "left turn" and "downgrade." Each road section in the Utah roadlog file may contain more than one horizontal curvature or vertical grade, and two ways of resolving this problem were considered: (1) using surrogate measures and (2) disaggregating road sections into smaller subsections. They are discussed as follows.

Three surrogate measures for horizontal curvature and three surrogate measures for vertical grade were devised to characterize the horizontal and vertical alignments of each road section. On a particular road section i with length ℓ_i (in miles), assume that along the length of the section there are K curved subsections, indexed by $k=1,2,...,K$, associated with it. Each subsection k has length $\ell_{i,k}$ and curvature $\theta_{i,k}$ (which could be positive or negative). Similarly, we assumed that there were G different vertical graded subsections associated with a road section, and each subsection had length $\ell_{i,g}$ and grade $\omega_{i,g}$ (which could be positive or negative), where $g=1,2,...,G$. These surrogate measures for a section i were defined as follows:

- Horizontal curvature change rate (CCR) and vertical grade change rate (GCR):

$$\mathrm{CCR}_i = \sum_{k=1}^{K-1} |\theta_{i,k+1} - \theta_{i,k}|$$
$$\mathrm{GCR}_i = \sum_{g=1}^{G-1} |\omega_{i,g+1} - \omega_{i,g}| \qquad (1)$$

  If there is only one curvature (or grade), CCR (or GCR) is defined as zero.

- Mean absolute horizontal curvature (MAC) and mean absolute vertical grade (MAG):

$$\mathrm{MAC}_i = \left(\sum_{k=1}^{K} \ell_{i,k} |\theta_{i,k}|\right)/\ell_i$$
$$\mathrm{MAG}_i = \left(\sum_{g=1}^{G} \ell_{i,g} |\omega_{i,g}|\right)/\ell_i \qquad (2)$$

- Maximum absolute horizontal curvature (MC) and maximum absolute vertical grade (MG):

$$MC_i = \max\{|\theta_{i,1}|, |\theta_{i,2}|, \ldots, |\theta_{i,K}|\} \qquad (3)$$
$$MG_i = \max\{|\omega_{i,1}|, |\omega_{i,2}|, \ldots, |\omega_{i,G}|\}$$

All of the surrogate curvature measures above are in the unit of degrees per 100-ft (30.5-m) arc, and surrogate grade measures in percent. Similar concepts were used in Joshua and Garber.[42]

Some modeling results using these surrogate measures were reported in Miaou, et al.[51] Limitations of these measures include the following: (1) it will be difficult for design engineers to incorporate these measures into their current practice directly; (2) these surrogate measures are hard to interpret in a design context; (3) these surrogate measures are not unique, i.e., different combinations of curves and grades can result in the same values; and (4) analyses of length of curve, length of grade, and continuous geometric design conditions are difficult, if not impossible, when these surrogate measures are used for the study of geometric design-accident relationships. In view of the above limitations, this report focuses on the second approach, i.e., to disaggregate those road sections with multiple curvatures and grades into smaller subsections in such a way that each subsection contains a unique set of horizontal curvature and vertical grade. Noted, however, that the second method would result in more short road sections than the first method. And because the locations of accidents are often estimated and not always accurate, accidents could be assigned to wrong road sections when road sections are too short (e.g., less than 0.05 mi or 0.08 km). In this study, we tested the effect of small road sections on model estimation (see chapter 4).

Illinois

In the Illinois roadlog file, each road section is homogeneous in terms of horizontal curvature, but not necessarily in terms of vertical grade. The approaching grade and leaving grades of each section are given for each road section. It is, however, not specified how many different grades are contained in a road section. It could happen that approaching grade is uphill (positive grade) while leaving grade is downhill (negative grade), or vice versa.

In the vehicle file, vehicles involved in accidents are categorized into 12 vehicle types (or body styles) including a "not stated" category. The truck category of interest includes: single unit truck (excluding pickups and vans), truck and tractor trailer. It is not clear how "bobtails" are categorized. However, based on the evaluation of 50 police accident reports from the Illinois State Department of Transportation, "bobtails" are probably categorized in a catch-all category called "other vehicles."

Unlike Utah, highway geometric design (including horizontal curvature and vertical grade), traffic, and other relevant information on each road segment are contained in one single roadlog file. That is, each road section in the roadlog file has already had a unique set of horizontal curvature. However, only 1987 roadlog data for Illinois are currently available in the HSIS. Therefore, to use the Illinois accident data from 1985 to 1987, one has to assume that the road characteristics did not change over the 3-year period.

Several issues arose during the preparation of Illinois data in an earlier study:[47]

- The Illinois roadlog file does not include information on lane width; instead information on <u>surface width</u> is included. A computational procedure has to be developed to calculate individual lane widths from the surface width. The procedure required information of road sections on median type, median width, number of lanes, pavement type, allowance for parking lanes, shoulder width, highway functional class, and other variables. The researchers' experience working with data on Illinois freeway and expressway suggested that although most of the computed lane widths were either 11, 12, or 13 ft (3.35, 3.66, or 3.96 m), quite a few road sections have lane width which varied from 6 to 24 ft (1.83 to 7.32 m), indicating some coding errors. However, it should be mentioned that HSRC has recently performed a similar calculation for two-lane arterial in Illinois and found that virtually all of the resulting lane widths fell between 10 and 13 ft (3.05 and 3.96 m) (personal communication with Mr. F. Council at HSRC).

- Horizontal curvature and vertical grade were recorded only for those road sections that were considered to be potentially "substandard." For example, horizontal curvature was only in the file for curves with degree of curve greater than 2.5 degrees, only in rural area, and not on Interstate roadways. Curves on Interstate are not included in the file since they are not considered potentially substandard.

- The road inventory recorded in the roadlog file is not complete. Approximately 4,000 accidents could not be merged with the roadlog file because the route number reported in the accident file does not exist in the roadlog file. Another 8,000 accidents can not be merged because no road section in the roadlog file has corresponding beginning and ending milepoints as reported in the accident file. We did not further investigate on which roadway classes these "unmatched" accidents occurred. It is expected that the majority occurred on local routes.

- Although it is indicated in Council and Williams that the roadlog file is updated annually, an examination of the data suggested that there was a fraction of road sections for which the AADT was recorded before 1985: 0.7 percent recorded in 1960's and 1970's, 7.3 percent between 1980 and 1984.[50]

The data were edited in the same manner as the Utah file. The total number of road sections finally selected for analysis was 2,348 for freeway and expressway in 1987. Only 15 out of these road sections were found to have curvature information, and all of them have curvature degree coded as greater than 2 degrees per 100-ft (30.5-m) arc. Most of the road sections considered are four-lane divided highways. For the vertical grade, two types of information were available for each road section--approach grade and leave grade. There were 128 (out of the total 2,348) road sections with either approach grade or leave grade greater than zero. Unfortunately, 69 out of these 128 graded sections either approach uphill and leave downhill or vice versa. That is, the approach grade and the leave grade of a particular road section have opposite signs.

AVAILABLE VARIABLES

Utah Data

For Utah, the variables that are relevant to this study and their current conditions are as follows:[49]

- Section length.
- Horizontal curvature: curvature degree, length of the curve (only for State system roadways).
- Vertical grade: grade percentage, direction of grade (however, the accident data are not specified by direction of grade).
- Lane width: (though 38 percent of the road sections are uncoded).
- Shoulder width and type: inside and outside shoulder widths per direction.
- Median type (and width): divided or undivided only.
- Number of lanes (both direction).
- Pavement conditions: service rating (87 percent of the road sections are uncoded).
- Speed limit: posted speed limit (79 percent of the road sections are uncoded); impact speed (50 percent coded as zero which could be a parked vehicle or an uncoded case).
- Annual average daily traffic (AADT): average AADT (but not available by time of day).
- Percent trucks: on-peak and off-peak truck percentages (not broken down by truck type).
- DUI: driver alcohol percent (BAC) (however, these data are potentially erroneous, and the associated truck exposure information is not available).[49]
- Lighting conditions: data on whether accidents occurred during night/dawn/dusk conditions (but not the associated truck exposure information).
- Weather: accidents that occurred under rainy/snowy/foggy conditions (but not the associated truck exposure information).

Currently, geometric and traffic data of interchanges and intersections are not available in HSIS for Utah. Although map documentation of interchanges on Interstate highways can be obtained for analysis purposes, one needs to match accident data to the map documentation manually.[49] In addition, Utah officials noted that there could be some inconsistencies in the ramp accident data. Another HSIS State--Minnesota--on the other hand, has a rather complete road inventory file just for intersections and interchanges.

Illinois Data

For Illinois, the variables that are relevant to this study and their current conditions are as follows:[50]

- Section length.
- Horizontal curvature: radius and deflection angle (only coded for "potentially substandard" road sections, i.e., road sections with horizontal curvature of approximately 2.5 degrees per 100-ft (30.5-m) arc or over).
- Vertical grade: approach and leave grades, grade percentage, direction of grade (however, the accident data are not specified by direction of grade).
- Surface width: lane width can be computed based on surface width (however, as indicated earlier, we experienced some problems with the computed lane width).
- Shoulder width and type: shoulder width per direction.
- Median type (and width): divided or undivided only.
- Number of lanes.
- Pavement conditions: surface type and surface condition rating.
- Speed limit: posted speed limit (47 percent of the road sections are coded as "unknown" due to the fact that the variable is coded only where speed limit changes).
- Annual average daily traffic (AADT): not available by time of day.
- Percent trucks: average truck percentages for each road section (not broken down by truck type), heavy commercial truck volume.
- DUI: driver alcohol percent (BAC) (these data are, however, provided for only 67 percent of drivers who were noted as having been tested, and the associated truck exposure information is not available).[50]
- Lighting conditions: data on whether accidents occurred during night/dawn/dusk conditions (but not the associated truck exposure information).

- Weather: accidents that occurred under rainy/snowy/foggy conditions (but not the associated truck exposure information).

Currently, geometric and traffic data of interchanges and intersections are not available for Illinois in the HSIS.

SUMMARY STATISTICS OF UTAH DATA

In this study, models relating truck accidents and highway geometric design variables are developed using Utah data for each of the following three roadway classes: rural Interstate, urban Interstate and freeway, and rural two-lane undivided arterial. The basic analysis time period used in this study is a 1-year period since there was a separate Utah road file for each year in the 5-year period. Thus, changes that occurred on the same road section, such as AADT and shoulder width, across time would be reflected in differences between the years. If nothing had changed on a road section between 1985 and 1989, identical inventory data would be used in each of the 5 years for analysis purposes.

In order for the established relationships to be representative, only those road sections with geometric design variables within "normal" ranges were considered. Table 2 shows the considered ranges of several geometric design variables. For example, the range for the absolute value of horizontal curvature (HC) was between 0 and 12 degrees per 100-ft (or 30.48-m) arc for rural Interstate highways. American Association of State Highway and Transportation Officials (AASHTO) "green book" was used as a guideline in determining these ranges.[52] For urban Interstate and freeway road sections, the number of lanes varied from 2 to 13 lanes. In this study, only road sections with four to eight lanes were considered. As a result, about 3 percent of the urban Interstate and freeway road sections were removed from analysis.

The total number of homogeneous road sections and lane-miles under consideration for the three roadway types during the 5-year period are:

- Rural Interstate: 8,263 sections, 14,731 lane-mi (23 570 lane-km).
- Urban Interstate and freeway: 2,810 sections, 3,889 lane-mi (6 222 lane-km).
- Rural two-lane undivided arterial: 13,634 sections, 9,211 lane-mi (14 738 lane-km).

Data for each year contains roughly one-fifth of the total sections and lane-miles above. General descriptive statistics of these road sections and the associated truck accident, truck exposure, traffic, and key highway geometric design variables are given at the end of this section in tables 3 to 5 for the three roadway types. The following is a discussion of the statistics summarized in these tables.

Table 2. Ranges of geometric design variables considered in this study.

Geometric Design Element	Roadway Class	AASHTO's Recommendation	Recorded Maximum	Considered Maximum	Associated Factors
Horizontal Curvature (degree/100 ft arc)	Rural Interstate Urban Interstate & Freeway Rural 2-Lane Undivided Arterial	≤ 5.75 (DS=60 mph; e=0.12) ≤ 5.75 (DS=60 mph; e=0.12) ≤ 22.75 (DS=30 mph; e=0.08)	19 9 108	12 9 30	Design Speed (DS), Spiral (Transition) Curve, Superelevation Rate (e)
Vertical Grade (percent)	Rural Interstate Urban Interstate & Freeway Rural 2-Lane Undivided Arterial	≤ 6% (DS=60 mph; t=mount.) ≤ 7% (DS=50 mph; t=mount.) ≤12% (DS=30 mph)	8% 12% 11%	8% 12% 11%	Design Speed (DS), Terrain, Length of Grade
Paved Inside Shoulder Width per Direction (ft)	Rural Interstate Urban Interstate & Freeway Rural 2-Lane Undivided Arterial	Ideal width = 12 ft (DDHV ≥ 250 trucks)	8 10 --	8 10 --	Directional Design Hourly Volume (DDHV), Horizontal Curvature, Number of Lanes
Paved or Stabilized Outside Shoulder Width per Direction (ft)	Rural Interstate Urban Interstate & Freeway Rural 2-Lane Undivided Arterial	Same as Above	10 12 23	10 12 12	Same as above

AASHTO: American Association of State Highway and Transportation Officials.

Accident Data

During the 5-year period, there were 1,643 large trucks involved in accidents on Utah rural Interstate, 1,904 on urban Interstate and freeway, and 789 on rural two-lane undivided arterial, regardless of truck configuration and accident severity type. With the total truck miles estimated to be 2,030 million truck miles (MTM), 1,044 MTM, and 694 MTM (1 MTM = 1.61 million truck kilometers), respectively, the overall truck accident involvement rate was 0.81, 1.82, and 1.14 truck involvements per MTM. A detailed description of these truck accidents stratified by roadway class, truck configuration, and accident severity types is given in the next section. These accidents occurred quite sporadically across road sections. For example, 86.19, 66.76, and 95.20 percent of the road sections had no truck involved in accidents being observed for the three roadway types, respectively, in a given year. While the maximum number of observed truck accidents on an individual road section was 8, 19, and 7, respectively, on these three roadway types. On the average, each road section in Utah had about 0.20, 0.68, and 0.06 trucks involved in accidents per year on these three roadway types.

Traffic and Truck Exposure Data

For each highway section, traffic and exposure data, including AADT, on-peak and off-peak truck percentages, and section length, are available from the roadlog file. For each road section i, truck exposure was computed as $v_i = 365 \times AADT_i \times (T\%_i/100) \times \ell_i$, where $T\%_i$ is the average percent of trucks determined from on-peak and off-peak truck percentages and ℓ_i is the section length in miles of section i. For rural highways, such as rural Interstate and rural two-lane arterial, on-peak and off-peak truck percentages are almost always given the same value for all road sections in the Utah roadlog file. For those road sections that on-peak and off-peak truck percentages are given different values, especially road sections located in urban areas, the truck percentage $T\%_i$ was computed as a weighted average of one-quarter of the on-peak percentage and three-quarters of the off-peak percentage. That is, $T\%_i = 0.25 \times$(on-peak percent trucks) + $0.75 \times$(off-peak percent trucks).

The percent truck data are currently not broken down by truck type. Table 5 shows the average vehicle distribution by highway functional class in 1989 for Utah, based on the relative areawide truck percentages provided in the Highway Performance Monitoring System.[46] For example, the table indicates that there were 5.9 and 19.1 percents of single-unit and combination trucks, respectively, on rural Interstate. When truck percentages broken down by these two truck types for an individual road section i were needed, we estimate the percentage of single-unit and combination trucks on section i as $T\%_i \times [5.9/(5.9+19.1)]$ and $T\%_i \times [19.1/(5.9+19.1)]$, respectively.

For comparison purposes, the relative percentages of vehicle travel by vehicle type from HPMS for different highway functional classes in Illinois are presented in table 6.

For a given highway supply (in lane-miles), AADT is typically used to indicate traffic condition or congestion level of a road section. Because the number of lanes varies from road section to road section, in this study "AADT per lane" was considered instead. This new variable represents the average density of vehicle flow on the road in an average day. Conceptually, the higher the vehicle flow density, the greater the chance for a truck to be involved in a conflicting position with other vehicles when negotiating its way through the road section.

Highway Geometric Data

For these three roadway classes, the section lengths vary from 0.01 to 7.77 mi, 0.01 to 3.02 mi, and 0.01 to 9.41 mi -- with an average of 0.45, 0.26, and 0.34 mi (1 mi =1.61 km). Because most of the road sections are coded as having 12-ft (3.66-m) lane width, we were unable to distinguish the effects of different lane widths on the truck accident rate in developing the models. In addition, the majority of the road sections on Interstate and freeway are divided. Some of the highway geometric variables used in the model development are as follows.

Shoulder width per direction:

Shoulder widths were recorded separately for inside (or left) and outside (or right) shoulders in a given direction. In this study, paved inside and outside shoulder widths were considered for Interstate and freeways, and stabilized (including paved) outside shoulder width was considered for rural two-lane undivided arterial. Also, a shoulder width of 12 ft (3.66 m) was considered to be an "ideal" shoulder width for both the inside and outside shoulders in that it practically adds an additional lane on each side of the road. Note that AASHTO's "green book" recommends that highways carrying large numbers of trucks should have usable shoulders at least 10 ft (3.05 m) and preferably 12 ft (3.66 m) wide. Furthermore, a variable called "deviation from the ideal shoulder width," which is the shoulder width short of the "ideal" shoulder width, was defined. Specifically, for road section i deviation from ideal shoulder width, denoted by SWD_i, was defined as $SWD_i = max\{0, 12 - SW_i\}$, where SW_i is the inside or outside shoulder width of section i. For Interstate and freeways, SW_i was the "paved" inside or outside shoulder width of section i, and for rural two-lane undivided arterial SW_i was the stabilized outside shoulder width of section i. Note that for rural Interstate and urban Interstate and freeway, because most of the road sections were recorded as having 10 ft (3.05 m) paved outside shoulder width, it was not possible to study the effects of paved outside shoulder width on truck accident rate.

Horizontal curvature and vertical grade

In this study, the absolute values of horizontal curvature (HC) and vertical grade (VG) on each homogeneous section were used as the explanatory variables. With these homogeneous sections, the maximum horizontal curvatures are 12, 9, and 30 degrees per 100-ft (30.5-m) arc for the three roadway types, respectively; while the corresponding maximum vertical grades are 8, 12, and 10 percent.

Length of original curve and length of original grade

It has been suggested in the literature that, for a fixed curvature degree, as the length of curve increases, the accident rate increases.[53] Also, as the length of grade increases to a point that can slow a truck to a speed significantly slower than the speed of the traffic stream (by, e.g., 10 mi/h or 16 km/h), the accident rate increases.[39] In order to test the effects of length of curve and length of grade on truck accident rate, two geometric design variables -- length of original curve (LHC) and length of original grade (LVG) -- were considered. As indicated earlier, each homogeneous curve or grade considered in this study may have been subdivided from a longer curve or grade for achieving total homogeneity. Thus, for each road section , these two explanatory variables were defined as the length of the original undivided curve or undivided grade to which this section belonged. In addition, these two variables were defined only for curves with horizontal curvature greater than 1 degree per 100-ft (30.48-m) arc and sections with grade greater than 2 percent. That is, these two variables were set equal to 0 if horizontal curvature is less than or equal to 1 degree or if vertical grade is less than or equal to 2 percent. This definition was based on an assumption that the length of a mild curve or grade has no aggravated effect on truck accident involvements.

TRUCK ACCIDENT FREQUENCY AND INVOLVEMENT RATE IN UTAH

This section highlights significant characteristics of Utah truck accidents from 1985 to 1989 for three roadway types: rural Interstate, urban Interstate and freeway, and rural two-lane undivided arterial.

For each roadway type, table 8 shows the number of trucks involved in accidents and the associated accident involvement rates by truck configuration (single-unit trucks, combination trucks, and all trucks combined), and by accident severity type (fatal and incapacitating injury, non-incapacitating injury and possible injury, property damage only, and all accident severity types combined). Total number of road sections and the associated lane-miles and truck exposure in the 5-year period are also included in the footnotes for the table. In general, the overall accident involvement rates for the three roadway types in table 8 are consistent with (but slightly lower than) those obtained in Harwood, et al. for California, Illinois, and Michigan (see table 9 for comparison).[(2)] It was observed that combination trucks have consistently higher accident involvement rate in almost all accident categories specified in table 8.

Tables 10, 11, and 12 give the relative percentages of trucks involved in accidents under different lighting and weather conditions. The accident involvement rates are not computable because the truck exposure data in the HSIS are not available by time of day and weather conditions. The first observation one can make from table 10 is that a significant percentage of these truck accidents were found to have occurred under dark, i.e., night, dawn, and dusk, conditions. For example, over 45 percent of the combination-truck accidents on rural highways have occurred under dark conditions. Furthermore, it was observed that a consistently higher proportion of combination trucks were involved in accidents under dark conditions than single-unit trucks. Most likely, this is a result of higher percentage of truck travel incurred at nighttime by the combination trucks. From tables 11 and 12, one can observe that about 15 percent to 20 percent of the combination-truck accidents occurred under either rainy (table 11) or snowy (table 12) conditions.

SUMMARY

In general, the HSIS was found to be a rich and well-prepared data base in terms of its ability to provide accident, vehicle, driver, and road information for highway safety research. The researchers' experience working with Utah data indicated that the truck accident involvement rates derived from the HSIS for four roadway types are generally consistent with other studies. However, the data base is currently limited in providing the following information:

- Detailed truck exposure data, e.g., by truck type, time of day, and weather conditions.
- Roadside design data, e.g., sideslope, ditch width, and roadside objects.
- Superelevation of curved road sections.
- Data on interchanges and intersections.

At present, because different highway geometric variables are collected and, occasionally, different definitions are used for variables with the same names collected by individual States, the merging of data from different States is expected to be difficult.

Table 3. Descriptive statistics of road sections for rural Interstate in Utah: 8,263 homogeneous sections.

Accidents, Exposure, Explanatory Variables	Mean	Standard Deviation	Skewness	Minimum	Maximum	Percent of Zeros
Truck accident Involvements	0.1988	0.5841	4.0755	0	8.00	86.19
Truck Exposure (Million Truck Miles)	0.2457	0.4161	4.4310	0.0008	5.03	0
AADT (in 1000's of vehicles)	7.8094	8.2459	3.6121	1.4000	71.41	0
AADT/Lane (in 1000's)	1.8030	1.4950	2.7380	0.3500	12.04	0
Horizontal Curvature (degrees/100-ft arc)	1.0047	1.9668	2.3752	0	12.00	67.46
Length of Original Curve (mi)	0.0475	0.1306	3.5000	0	0.96	80.85
Vertical Grade (percent)	2.1423	1.5851	0.8481	0	8.00	19.92
Length of Original Grade (percent)	0.2148	0.4845	2.9466	0	3.85	74.34
Paved Inside Shoulder Width per Direction (ft)	3.8393	0.9191	-3.0905	0	8.00	4.85
Paved Outside Shoulder Width per Direction (ft)	10.0000	0	0	10.0000	10.00	0
Percent Trucks (percent)	24.1270	8.0715	0.2993	7.0000	57.00	0
Section Length (mi)	0.4479	0.7232	4.1126	0.0100	7.77	0

Notes: (1) All sections have 12 ft (3.66 m) lane width; and (2) the number of sections that have 2, 3, 4, 6, and 9 lanes is 253, 148, 7341, 511, and 10, respectively.

(1 ft = 0.3048 m, 1 mi = 1.61 km)

Table 4. Descriptive statistics of road sections for urban Interstate and freeway in Utah: 2,810 homogeneous sections.

Accidents, Exposure, Explanatory Variables	Mean	Standard Deviation	Skewness	Minimum	Maximum	Percent of Zeros
Truck accident Involvements	0.6776	1.4809	4.8660	0	19.00	66.76
Truck Exposure (Million Truck Miles)	0.3714	0.5216	4.1885	0.0015	6.03	0
AADT (in 1000's of vehicles)	44.3653	29.8843	1.2142	2.4800	157.73	0
AADT/Lane (in 1000's)	7.6940	4.4250	1.0690	0.6200	21.67	0
Horizontal Curvature (degrees/100-ft arc)	0.8399	1.5456	2.3275	0	9.00	67.19
Length of Original Curve (mi)	0.0871	0.1983	2.4990	0	1.02	78.22
Vertical Grade (percent)	2.0740	1.5024	3.1433	0	12.00	14.63
Length of Original Grade (percent)	0.0615	0.1889	4.0469	0	1.25	85.77
Paved Inside Shoulder Width per Direction (ft)	3.9908	0.6654	2.2296	0	10.00	1.25
Percent Trucks (percent)	11.0401	5.7830	0.9932	1.0000	35.00	0
Section Length (mi)	0.2555	0.3171	3.6919	0.0100	3.02	0

Notes: (1) All sections have 12 ft (3.66 m) lane width; (2) the number of sections that have 4, 5, 6, 7, and 8 lanes is 893, 26, 1761, 4, and 126, respectively; and (3) 2,679 (or 95.3 percent of) road sections have a paved outside shoulder width of 10 ft (3.05 m) in each direction.

(1 ft = 0.3048 m, 1 mi = 1.61 km)

Table 5. Descriptive statistics of road sections for rural two-lane undivided arterial in Utah: 13,634 homogeneous sections.

Accidents, Exposure, Explanatory Variables	Mean	Standard Deviation	Skewness	Minimum	Maximum	Percent of Zeros
Truck accident Involvements	0.0579	0.2853	6.9576	0	7.00	95.20
Truck Exposure (Million Truck Miles)	0.0509	0.1301	5.6957	0.0004	1.95	0
AADT (in 1000's of vehicles)	2.5139	1.9250	1.9003	0.2150	16.20	0
AADT/Lane (in 1000's)	1.2570	0.9625	1.9000	0.1075	8.10	0
Horizontal Curvature (degrees/100-ft arc)	2.3846	4.4672	2.6105	0	30.00	66.73
Length of Original Curve (mi)	0.0446	0.0893	3.0307	0	0.84	67.05
Vertical Grade (percent)	2.7003	1.6163	1.6701	0	11.00	5.03
Length of Original Grade (percent)	0.2143	0.5510	4.9734	0	6.42	69.91
Paved Outside Shoulder Width per Direction (ft)	4.2431	2.8545	0.3143	0	12.00	8.59
Percent Trucks (percent)	17.6197	10.8279	0.9721	1.0000	56.00	0
Section Length (mi)	0.3378	0.6840	4.5239	0.0100	9.41	0

Notes: The number of road sections that have 11- and 12-ft (3.35- and 3.66-m) lane widths is 2 and 13,632, respectively.

(1 ft = 0.3048 m, 1 mi = 1.61 km)

Table 6. Distribution of annual miles of travel by vehicle type and functional class for Utah, 1989.

	Passenger Car	Two-Axle 4-Tire Truck	Other Single-Unit Truck	Combination Truck	Other	Total
RURAL						
Interstate	57.3%	17.4%	5.9%	19.1%	0.3%	100%
Other Principal Arterial	54.6%	28.3%	4.6%	12.4%	0.1%	100%
Minor Arterial	56.9%	33.0%	4.2%	5.7%	0.2%	100%
Major Collector	62.2%	31.1%	3.1%	3.4%	0.2%	100%
Minor Collector	58.9%	33.8%	4.0%	3.3%	0.0%	100%
Local	50.4%	48.4%	1.2%	0.0%	0.0%	100%
URBAN						
Interstate	66.2%	23.0%	3.5%	7.1%	0.2%	100%
Other Freeway Expressway	65.1%	26.4%	3.5%	4.9%	0.1%	100%
Other Principal Arterial	85.2%	12.7%	1.1%	0.9%	0.1%	100%
Minor Arterial	91.9%	6.2%	1.0%	0.8%	0.1%	100%
Collector	91.3%	8.6%	0.1%	0.0%	0.0%	100%
Local	89.5%	10.5%	0.0%	0.0%	0.0%	100%

Note: Data are collected for the Highway Performance Monitoring System and are provided by Utah State Department of Transportation.

Table 7. Distribution of annual miles of travel by vehicle type and functional class for Illinois, 1989.

	Passenger Car and 4-Tire Vehicle	Other Single-Unit Truck	Combination Truck	Other	Total
RURAL					
Interstate	75.0%	3.0%	22.0%	0.0%	100%
Other Principal Arterial	88.0%	3.0%	8.0%	1.0%	100%
Minor Arterial	90.0%	3.0%	7.0%	0.0%	100%
Major Collector	93.0%	3.0%	4.0%	0.0%	100%
Minor Collector	95.0%	3.0%	1.0%	1.0%	100%
Local	93.0%	5.0%	1.0%	1.0%	100%
URBAN					
Interstate	91.0%	3.0%	6.0%	0.0%	100%
Other Freeway Expressway	94.0%	2.0%	4.0%	0.0%	100%
Other Principal Arterial	95.0%	3.0%	2.0%	0.0%	100%
Minor Arterial	96.0%	2.0%	2.0%	0.0%	100%
Collector	97.0%	1.0%	1.0%	1.0%	100%
Local	97.0%	1.0%	1.0%	1.0%	100%

Note: Data are collected for the Highway Performance Monitoring System and are provided by Illinois State Department of Transportation.

Table 8. Number of large trucks involved in accidents and involvement rates for three roadway types in Utah: 1985-1989.

Truck Type →	Single-Unit Trucks				Combination Trucks				All Trucks			
Severity Type →	K&A	B&C	PDO	Total	K&A	B&C	PDO	Total	K&A	B&C	PDO	Total
Roadway Type ↓												
Rural Interstate	23 0.05	49 0.10	107 0.22	179 0.37	240 0.15	309 0.20	915 0.59	1464 0.94	263 0.13	358 0.18	1022 0.50	1643 0.81
Urban Interstate & Freeway	49 0.14	137 0.39	344 0.99	530 1.52	142 0.20	250 0.36	982 1.41	1374 1.98	191 0.18	387 0.37	1326 1.27	1904 1.82
Rural Two-Lane Undivided Arterial	24 0.11	31 0.14	93 0.42	148 0.68	120 0.25	117 0.25	404 0.85	641 1.34	144 0.21	148 0.21	497 0.72	789 1.14

Notes:
1. Accident severity type: K=Fatal, A=Incapacitating Injury, B=Non-Incapacitating Injury, C=Possible Injury, and PDO= Property Damage Only.
2. Accident involvement rates are presented in trucks per million truck miles.
3. Bobtails, i.e., tractors with no trailers, were not included in the table.
4. Number of sampled highway sections and total lane-miles per 5-year period:
   Rural Interstate: 8,263 sections, 14,730.9 lane-miles;
   Urban Interstate & Freeway: 2,810 sections, 3,888.9 lane-miles;
   Rural Two-Lane Undivided Arterial: 13,634 sections, 9,211.0 lane-miles.
5. Vehicle miles of travel (VMT) per 5-year period by truck type:
   Rural Interstate: Single Units = 479 million, Combinations = 1,551 million;
   Urban Interstate & Freeway: Single Units = 348 million, Combinations = 696 million;
   Rural Two-Lane Undivided Arterial: Single Units = 217 million, Combinations = 477 million.

(1 mi = 1.61 km)

Table 9. Comparison of truck accident involvement rates (in number of trucks per million truck miles) in four States.[2]

Roadway Type	California	Illinois	Michigan	Utah
Rural Interstate	0.53	0.46	1.18	0.81
Urban Interstate & Freeway	1.59	5.82	2.80	1.82
Rural Two-Lane Undivided Arterial	1.73	3.13	2.22	1.14

(1 mi = 1.61 km)

Table 10. Number and percentages of large trucks involved in accidents occurred under dark (night/dawn/dusk) conditions.

Truck Type →	Single-Unit Trucks				Combination Trucks				All Trucks			
Severity Type →	K&A	B&C	PDO	Total	K&A	B&C	PDO	Total	K&A	B&C	PDO	Total
Roadway Type: ↓												
Rural Interstate	5 21.7	14 28.6	30 28.0	49 27.4	112 46.7	141 45.6	443 48.4	696 47.5	117 44.5	155 43.3	473 46.3	745 45.3
Urban Interstate & Freeway	8 16.3	19 13.9	36 10.5	63 11.9	46 32.4	68 27.2	207 21.1	321 23.4	54 28.3	87 22.5	243 18.3	384 20.2
Rural Two-Lane Undivided Arterial	5 20.8	8 25.8	28 30.1	41 27.7	53 44.2	53 45.3	190 47.0	296 46.2	58 40.3	61 41.2	218 43.9	337 42.7

Table 11. Number and percentages of large trucks involved in accidents occurred under rainy conditions.

Truck Type →	Single-Unit Trucks				Combination Trucks				All Trucks			
Severity Type →	K&A	B&C	PDO	Total	K&A	B&C	PDO	Total	K&A	B&C	PDO	Total
Roadway Type: ↓												
Rural Interstate	0 0.0	1 2.0	6 5.6	7 3.9	14 5.8	14 4.5	46 5.0	74 5.1	14 5.3	15 4.2	52 5.1	60 4.9
Urban Interstate & Freeway	2 4.1	7 5.1	14 4.1	23 4.3	11 7.8	11 4.4	45 4.6	67 4.9	13 6.8	18 4.7	59 4.5	90 4.7
Rural Two-Lane Undivided Arterial	1 4.2	1 3.2	3 3.2	5 3.4	7 5.8	9 7.7	10 2.5	26 4.1	8 5.6	10 6.8	13 2.6	31 3.9

Table 12. Number and percentages of large trucks involved in accidents occurred under snowy conditions.

Truck Type →	Single-Unit Trucks				Combination Trucks				All Trucks			
Severity Type →	K&A	B&C	PDO	Total	K&A	B&C	PDO	Total	K&A	B&C	PDO	Total
Roadway Type: ↓												
Rural Interstate	3 13.0	8 16.3	22 20.6	33 18.4	24 10.0	46 14.9	151 16.5	221 15.1	27 10.3	54 15.1	173 16.9	254 15.5
Urban Interstate & Freeway	3 6.0	16 11.7	47 13.7	66 12.5	8 5.6	34 13.6	123 12.5	165 12.0	11 5.8	50 12.9	170 12.8	231 12.1
Rural Two-Lane Undivided Arterial	2 8.3	2 6.5	10 10.8	14 9.5	12 10.0	14 12.0	43 10.6	69 10.8	14 9.7	16 10.8	53 10.7	83 10.5

4. A POISSON REGRESSION MODEL

ISSUES IN DEVELOPING ACCIDENT MODELS

As mentioned earlier, vehicle accidents are complex processes involving the interactions of many factors: the road, the traffic, the driver, the vehicle, and the environment (e.g., weather and lighting conditions). Because some of the interacting factors are qualitative and stochastic in nature, e.g., drivers' behavior and weather conditions, the actual relationships between vehicle accidents and these factors are inevitably empirical and statistical. Previous studies indicated that totally controlled statistical experiment designs are extremely difficult to conduct under this situation.

The problems a researcher may face in establishing empirical relationships between vehicle accidents and highway geometric design have several important characteristics. They are summarized as follows.

1. **The occurrences of vehicle accidents are sporadic random events which are probabilistic in nature:** Given a road section, the number of vehicles involved in accidents during a period of time is a random variable taking nonnegative integer values (0,1,2,...), each of which has some probability of being observed, depending on the total vehicle exposure and accident involvement rate of the section during the period. The magnitudes of accident occurrence for passenger cars and combination trucks in the U.S. are in the neighborhood of 6.33 and 2.87 per million vehicle miles traveled, respectively.(1) On the other hand, road sections are usually delineated in an order of less than 1 mi (1.61 km) for ensuring that each individual road section is relatively homogeneous in geometric features and traffic conditions. Therefore, for a period of several years, the accumulated vehicle mileage on each road section is usually not large enough to observe a sufficient number of accidents (except those sections in large urban core areas, which carry very high vehicle volume on a daily basis). In most studies of this kind, especially truck related safety studies, the analyst is faced with a problem of dealing with a large number of road sections that have no accidents during the observed period. Zegeer, et al.'s study is a good example, in which 55.7 percent of the 10,900 curved road sections they studied had no vehicle accidents occurring in a 5-year period.(53) An even higher percentage of "no accident" road sections could be expected if only accidents involving trucks were studied.

2. **Road sections differ from each other not just in geometric features and traffic conditions, but also in the amount of vehicle exposure:** Because the amount of vehicle travel (or exposure) varies from one road section to another resulting from the differences in section lengths and/or vehicle volumes, the probability of observing the same number of vehicle accident involvements differs by road section even if the road sections are otherwise identical. For instance, for two road sections having identical geometric features and traffic conditions but with different amounts of vehicle exposure, the probability of observing no accidents is lower on the road section with high vehicle exposure than the one with low vehicle exposure.

3. **Vehicle accidents are complex interactions involving many factors: the road, the traffic, the driver, the vehicle, and the environment (e.g., weather and lighting conditions):** Although

the focus of this study is on establishing relationships between truck accidents and key highway geometric design elements, there are many other important factors that may affect accident involvement rate of a road section at the same time. The following is an example list of these potential factors:

Road:

- Horizontal curvature: curvature degree, length of curve, superelevation.
- Vertical grade: grade percentage, length of grade, direction of grade (upgrade or downgrade).
- Sight distance.
- Lane width, number of lanes, bridge width.
- Shoulder width and type: paved width, unpaved width.
- Median type and width/height: positive barriers, grass/trees.
- Roadside design: sideslope, ditch width, clear-zone width, number of entries and exits along the road section, guardrail, trees, utility poles, roadway lighting.
- Pavement surface conditions and edge drop: pavement type (or friction), service rating.
- Construction activity: whether the accident occurred when the road section was under construction.

Traffic:

- Annual average daily traffic (AADT): AADT by time of day, traffic density (e.g., AADT per lane) by time of day, percentage of trucks in the traffic stream.
- Vehicle speed: posted speed limit, top 25 percentile average vehicle speed, difference in mean speed between trucks and non-trucks.

Drivers:

- Cohort: average percent of (or truck miles incurred by) inexperienced (or young) truck drivers on the road.
- DUI: average percent of (or vehicle miles incurred by) drivers who were under the influences of drugs or alcohol.
- Driver factors: fatigue, judgement ability, mood, knowledge, physical and medical conditions, etc.

Vehicle:

- Vehicle type and performance: the composition of single and combination trucks (e.g., tractor semitrailer, truck trailer, twin trailer, ...) on the road.
- Vehicle age: average age of the trucks on the road.

Environment:

- Lighting conditions: percent of truck travel incurred under different lighting conditions.

- Weather: percent of truck travel incurred under wet (e.g., rainy and snowy) conditions, number of foggy days.
- Police enforcement level.

In empirical analysis, many of the variables listed above (or their proxy variables) will probably never be available for each individual road section. Therefore, in developing the models, one should recognize the fact that, no matter how many explanatory variables one manages to include, there are always some variables which will be excluded, especially those qualitative types of variables. It should be emphasized here that many vehicle accidents are directly or indirectly related to driver factors, and these driver factors are not likely to be found in any data base. That is, one will face, to some extent, the so-called "omitted variable" problem in developing the models.[54] In vehicle accident studies, this problem has traditionally been alleviated by developing separate models for different roadway classes and for different vehicle types when the data permit such an analysis, with the hope that these omitted variables will be roughly constant for a particular vehicle type within each roadway class. However, in truck accident studies, a limited number of truck accidents and a limited amount of truck exposure data may prohibit researchers from breaking down accident analysis by both roadway class and truck types.

In view of this inevitable omitted variable problem, when any geometric design effect is discussed, we have in mind the average observed effect, which includes the collective influences of all the interacting effects. This includes the influences of interacting factors such as driver's factors, vehicle speed, weather, etc. Thus, the geometric design effects we estimate are actually conditional on the omitted variables. That is, the effects of the same highway geometric design on vehicle accidents would be different if some of the omitted variables change. For example, changes in vehicle performance, socioeconomic, legislative, and law enforcement conditions over the years would change the geometric design effects on vehicle accidents even if nothing is done to the roads. For this reason, the analysts should always be careful in interpreting the estimated effects, be conscious of any potential bias, and be cautious in using the effects derived from one area for other areas.

4. **Vehicle accident data and exposure data are both subject to sampling and nonsampling errors:** Although there are several existing accident data bases, they are all subject to some degree of deficiency in a number of areas, including questionable quality.[55] One of the potential problems is under-reporting. In reality, not all accidents are reported or recorded, especially minor accidents. Also, the location of an accident is oftentimes estimated, and sometimes it is assigned to the nearest milepost of the route where it occurred. Therefore, assigning vehicle accidents to very short road sections is more susceptible to locational error than assigning to longer road sections. Vehicle exposure data, on the other hand, come primarily from FHWA's Highway Performance Monitoring System (HPMS), a highway sampling system statistically designed to obtain physical, traffic, and operational information on national highways from a small portion of selected highway sections.[56] Both vehicle volume (or AADT) and percent vehicles are subject to sampling error (e.g., daily, day-of-week, seasonal variations) and nonsampling error (e.g., vehicle axle counting and vehicle classification errors) from automatic vehicle recording machines.

In view of the characteristics of the problem described above, a potential model for establishing empirical relationships between truck accidents and highway geometric design ought to be a probabilistic model capable of:

- Addressing truck accident questions in terms of both accident involvement rate and accident probability.
- Predicting "nonnegative" accident involvement rates.
- Taking into account the differences in truck exposure across road sections.
- Giving proper weights probabilistically to a great portion of road sections with no observed truck accidents, depending on their truck exposure as well as other explanatory variables.
- Providing inferential statistics that allow the evaluation of model uncertainties due to the uncertainties of truck exposure data and possible omitted variables.

Whenever data permits, separate models should be developed for different roadway classes, truck types, and accident severity types to help alleviate the omitted variable problem.

MODELS USED IN PREVIOUS STUDIES

The empirical relationship between vehicle accidents and important highway geometric design elements, such as horizontal curvature, vertical grade, lane width, and shoulder width, has been addressed in numerous previous studies. Unfortunately, most of the studies did not distinguish vehicle accidents between trucks and other vehicles.[1,25]

Multiple Linear Regression

The relationships between vehicle accidents and highway geometric design have typically been established using the conventional multiple linear regression models, e.g., studies in NCHRP Report 197 and Zegeer, et al.[53] There are, however, several statistical properties of the conventional multiple linear regressions that are considered undesirable in establishing the relationships. These undesirable properties relate mainly to its underlying distributional assumption (normal distribution). The following are some examples:

- For a given road section, the number of vehicles involved in accidents during a period of time are random discrete events that take nonnegative integer values (0,1,2 ...). The use of a continuous distribution, like normal distribution, to model the accident event is at best an approximation to a truly discrete process.
- The occurrences of vehicle accidents are sporadic events. Zegeer, et al.'s study is a good example, in which 55.7 percent of the 10,900 curved road sections they studied had no vehicle accidents over a 5-year period.[53] This suggests that for a

period of several years most of the road sections considered would have a much higher probability of being observed with no accidents than with more than one accident. In other words, the underlying distribution of the occurrences of vehicle accidents on most of the road sections is positively (or rightly) skewed. The normal distribution is not a good approximation under this condition.

- Some inferential assumptions of multiple linear regression are probably too restrictive for this type of study, e.g., the residuals of the model are assumed to be uncorrelated with the explanatory variables.

- Other limitations include: (1) for some types of mathematical formulations, the models may occasionally predict negative accident involvement rates, and (2) the models do not provide a clear linkage between accident rate and accident probability. That is, given an accident rate, it is difficult to compute the probability of observing "y" vehicles involved in accidents on a particular road section during a period of time from the models.

Poisson Regression

In contrast to multiple linear regression models, Poisson regression models are widely used for modeling accident and mortality data in epidemiology. It is only in a recent study by Joshua and Garber that the model was introduced to establish the relationships between truck accidents and highway geometric design.[42,43]

The Poisson model used by Joshua and Garber assumed that the total number of truck accidents on a road section i during a 1-year period, y_i, follows a Poisson distribution. In addition, the expected number of truck accident involvements, $E(y_i)$, is related to k traffic and geometric variables, $\underline{x}_i=(x_{i1}=1,x_{i2},\ldots,x_{ik})'$, in the following form:

$$E(y_i)=f(\underline{x}_i;\underline{\beta})=\beta_1\left(\prod_{j=2}^{k}x_{ij}^{\beta_j}\right) \tag{4}$$

where $\underline{\beta}=(\beta_1,\beta_2,\ldots,\beta_k)'$ are model parameters to be estimated. In other words, y_i is postulated to follow a Poisson distribution with both mean and variance equal to $f(\underline{x}_i;\underline{\beta})$. Unlike the multiple linear regression model, this postulation suggests that the variance of y_i involves both the explanatory variables $\underline{x}_i$ and unknown parameters $\underline{\beta}$. Joshua and Garber reported that model parameters were estimated by a weighted least squares (WLS) procedure. It is, however, not clear from the report exactly how the WLS was carried out. Note that possible WLS methods include the weighting by vehicle travel as that used in Zegeer, et al., the iterative reweighted least squares (IRWLS) described in Carroll and Ruppert, and the quasi-likelihood method introduced in McCullagh and Nelder.[53,57,58]

It should be noticed from this particular formulation of Joshua and Garber that truck exposure in a year on a particular section i, v_i, computed as $v_i = 365 \times AADT_i \times (T\%_i/100) \times \ell_i$, where $T\%_i$ is the truck percentage and ℓ_i is the length of section i, is not explicitly represented in the model, but rather AADT, percent trucks, and section length were each included as a candidate explanatory variable, i.e., x's in eq (4). This particular formulation has some inherent limitations, which can be better illustrated by using one of the final models they developed.

1. The expected number of truck accidents in a year for each highway section of a given highway environment is estimated to be:

$$TINVOL = 0.015237 \times (SCR)^{0.0577} \times (AADT)^{0.5024} \times (TPERCENT)^{0.5731} \tag{5}$$

where SCR (slope change rate) is a surrogate vertical grade measure, and TPERCENT is the percent of trucks.[42] One obvious limitation with the model is that it will always give a small prediction of truck accidents for relatively level highway sections (i.e., SCR $\approx$ 0) no matter what other variables are being included in the model. In other words, regardless of the AADT and the percent of trucks, as long as SCR is rather small, the expected number of accidents based on the model would be small.

2. Another limitation, which relates to the truck exposure data, becomes obvious when deriving an accident rate, λ, from the model. That is:

$$\lambda = \frac{TINVOL}{365 \times AADT \times (TPERCENT/100) \times SECLEN} \tag{6}$$

$$= (0.015237 \times 100/365) \times (SCR)^{0.0577} \times (AADT)^{-0.4976} \times (TPERCENT)^{-0.4269} \times (SECLEN)^{-1}$$

where SECLEN represents section length in miles. Contrary to what one would usually expect, the model suggests that increases in AADT or section length, while holding other variables constant, would reduce the truck accident rate.

Although different explanatory variables were identified in modeling the other two highway environments in Joshua and Garber, similar limitations appeared in all three final models.

While the proposed model in our study was also in the Poisson context, it had a different model structure as that used in Joshua and Garber, which does not have the two limitations observed in the Joshua and Garber model. Furthermore, our proposed model addressed the consequences of a general limitation of using the Poisson regression model. That is, the variance of the data is restrained to be equal to the mean. This is well-known in the statistical literature.[59,60]

The remainder of the chapter is organized as follows. First, a Poisson regression model for establishing the relationships between truck accidents and highway geometric design variables is presented. Second, model results for three roadway types in Utah are discussed. Finally, directions for future work are suggested.

THE PROPOSED POISSON REGRESSION MODEL

The proposed framework makes use of the Poisson regression models. In this model, the number of trucks involved in accidents on each road section over a period of time is assumed to be Poisson distributed, and the Poisson rate is related to highway geometric, traffic, and other potential explanatory variables by a loglinear function. The specific truck safety questions this proposed model framework is intended to address include:

- Given a section of highway, how safe is it for large trucks in terms of both accident involvement rate and accident probability?
- Given a set of highway geometric design elements, which elements have relatively more impact on the safety performance of large trucks?
- What reduction in large truck accident involvement rates can be expected from various improvements in highway geometric design?

Although the discussion below does not distinguish accidents by truck type and accident severity, in principal, the overall framework could be used to any truck type and accident severity type of interest, provided that there are enough accident data and that truck exposure by truck type is available.

Model Formulation

Consider a set of n highway sections of a particular roadway type, say, rural Interstate. Let Y_i be a random variable representing the number of trucks involved in accidents on highway section i during a period of, say, 1 year where $i=1,2,...,n$. Further, assume that the amount of truck travel or truck exposure on this highway section, V_i, is also a random variable, estimated through a highway sampling system, such as Highway Performance Monitoring System.[56] Associated with each highway section i, there is a $k \times 1$ vector of explanatory variables (or independent variables, or covariates), denoted by $\underline{x}_i = (x_{i1}, x_{i2}, ... ,x_{ik})'$, describing its geometric characteristics, traffic conditions, and other relevant attributes. Note that without loss of generality, x_{i1} is defined as a dummy variable equal to one for all i (i.e., $x_{i1}=1$). Some of the variables can be 0,1 dummy variables, indicating the presence or absence of a condition. Given V_i and $\underline{x}_i$, truck accident involvements Y_i, $i=1,2,...,n$, are postulated to be independent, and each is Poisson distributed as:

$$p(Y_i=y_i|\Lambda_i=\lambda_i,V_i=v_i,\underline{x}_i)=\frac{(\lambda_i v_i)^{y_i}e^{-\lambda_i v_i}}{y_i!}, \qquad (i=1,2,...,n;\ y_i=0,1,2,...) \tag{7}$$

where Λ_i (>0) is the truck accident involvement rate on highway section i, and it is expected to vary from one highway section to another, depending on its explanatory variables $\underline{x}_i$. For each highway section i, the Poisson model implies that the conditional mean is equal to the conditional variance:

$$E(Y_i|\Lambda_i=\lambda_i,V_i=v_i,\underline{x}_i)=Var(Y_i|\Lambda_i=\lambda_i,V_i=v_i,\underline{x}_i)=\lambda_i v_i \tag{8}$$

and is proportional to truck exposure v_i. The definition and probabilistic properties of the Poisson process are well-known and will not be repeated here.[61]

To establish a relationship between truck accident involvement rate and highway geometric, traffic, and other variables, the following exponential form rate function is used in this study:

$$\Lambda_i=\exp(\underline{x}_i'\underline{\beta}+\varepsilon_i)=\exp(\underline{x}_i'\underline{\beta})\exp(\varepsilon_i)=\exp(\sum_{j=1}^{k}x_{ij}\beta_j)\exp(\varepsilon_i)=\lambda_i\exp(\varepsilon_i) \tag{9}$$

where $\underline{\beta}=(\beta_1,\beta_2,...,\beta_k)'$ is a $k \times 1$ parameter vector, and ε_i is a specification error due, for instance, to omitted variables which are independent of $\underline{x}_i$. Note that higher order and interaction terms of explanatory variables can be included in the equation without difficulties whenever appropriate. The specification error, ε_i, admits the fact that the functional relationship is at best an approximation to the true relationship. This particular loglinear relationship, which ensures that the involvement rate is always nonnegative, has been widely employed in statistical studies and found to be quite flexible in fitting different types of count data.[61, 62, 60, 63]

In cases where $\underline{x}_i$ and V_i are given with no (or negligible) uncertainties and Λ_i is assumed to be a constant (i.e., $\varepsilon_i=0$, for all i), then eq (7) becomes a classical Poisson regression model. However, if not so, the effects of the uncertainties in V_i and Λ_i create extra variations (or overdispersion) in the Poisson model.[60,64] In other words, the variance of the data will be greater than what the Poisson model predicted. The consequences of ignoring the overdispersion in the

Poisson regression are that consistent estimates of the regression parameters, $\underline{\beta}$, under the classical Poisson model, such as maximum likelihood estimates (MLE), are still consistent; however, the variances of the estimated parameters would tend to be underestimated. Specifically, the actual uncertainties of the estimated parameters will be higher than those indicated through a classical Poisson model.[59,65]

Throughout this chapter, the classical Poisson regression is used. That is, the model assumes that truck exposure V_i and explanatory variables $\underline{x}_i$ are observed without error and truck accident involvement rate, Λ_i, is a constant for each road section i. The potential underestimation of parameter variance, due to overdispersion in the Poisson regression model, will be adjusted using an estimate of overdispersion suggested by Wedderburn.[66] Furthermore, for ease of exposition, we will let $\Lambda_i=\lambda_i$ and $V_i=v_i$ and rewrite eqs (7) and (9) as:

$$p(y_i) = \frac{(\lambda_i v_i)^{y_i} e^{-\lambda_i v_i}}{y_i!}, \qquad \text{where} \quad \lambda_i = e^{\underline{x}_i' \underline{\beta}} = e^{\sum_{j=1}^{k} x_{ij}\beta_j} \tag{10}$$

This classic Poisson regression model assumes that Y_i are Poisson distributed with mean μ_i ($=\lambda_i v_i$), and the mean (or the expected number of trucks involved in accidents) μ_i is proportional to truck travel v_i. This model also assumes an exponential rate function, $\lambda_i=E(Y_i)/v_i=\exp(\underline{x}_i'\underline{\beta})$, which ensures that accident involvement rate is always nonnegative.

Based on the model, the variance, $Var(Y_i)$, and coefficient of skewness, $skew(Y_i)$, of the underlying distribution of Y_i are μ_i and $\mu_i^{-1/2}$, respectively. The variance, $Var(Y_i)$, which is equal to the mean μ_i, depends on its rate function and, thus, involves unknown regression parameters. (Recall that $\mu_i=\lambda_i v_i$ and λ_i is a function of unknown regression parameters). In addition, $Var(Y_i)$ grows linearly with truck exposure v_i. The model supposes a positive skewness coefficient which varies from road section to road section, depending on their means ($skew(Y_i)=\mu_i^{-1/2}$); as mean μ_i increases, either as a result of an increase in vehicle exposure v_i or an increase in the rate function λ_i, the skewness coefficient of the Poisson models decreases. This is a desirable property for two reasons: (1) when μ_i is reasonably large, one can expect the normal distribution to approximate the true probability distribution of Y_i well; and (2) for road sections with very small μ_i, the probability distribution of Y_i will be highly positively skewed.

The Poisson distributional assumption can be used to obtain tests and confidence statements about the estimated regression parameters and, unlike the conventional regression models which rely on normal assumption, this distribution can oftentimes be used to make reasonable probabilistic statements about Y_i.

Model Estimation and Diagnostic Checking

The Poisson model considered in this study assumes that the occurrences of truck accidents on different road sections and in different time periods are independent. Then, for the observed road sections and truck exposure, the likelihood function can be written as:

$$l(\underline{\lambda}) = \prod_i \frac{(\lambda_i v_i)^{y_i} e^{-\lambda_i v_i}}{y_i!} = e^{-\sum_i \lambda_i v_i} \left[\prod_i \frac{(\lambda_i v_i)^{y_i}}{y_i!} \right] \tag{11}$$

where $\underline{\lambda}$ is a vector representation of all λ_i's. The log-likelihood function is then:

$$L(\underline{\lambda}) = \sum_i \left[y_i \log(\lambda_i v_i) - \lambda_i v_i - \log(y_i!) \right] \tag{12}$$

Replacing λ_i with $\exp(\underline{x}_i'\underline{\beta})$, we have:

$$L(\underline{\beta}) = \sum_i \left[y_i(\underline{x}_i'\underline{\beta}) + y_i \log(v_i) - e^{\underline{x}_i'\underline{\beta}} v_i - \log(y_i!) \right] \tag{13}$$

To obtain the maximum likelihood estimates of regression parameters $\underline{\beta}$, denoted by $\hat{\underline{\beta}}$, the likelihood function is first differentiated as:

$$\frac{\partial L(\underline{\beta})}{\partial \beta_j} = \sum_i \left[y_i - e^{\underline{x}_i'\underline{\beta}} v_i \right] x_{ij} \qquad (j = 1, 2, \ldots, k) \tag{14}$$

and then set to zero. Since x_{i1} is a dummy variable equal to 1 for all i, the MLE requires that $\Sigma_i y_i = \Sigma_i v_i \exp(\underline{x}_i'\hat{\underline{\beta}})$. That is, the (estimated) expected total number of accident involvements, $\Sigma_i \hat{\mu}_i$, has to be equal to the observed total $\Sigma_i y_i$, where $\hat{\mu}_i = v_i \hat{\lambda}_i = v_i \exp(\underline{x}_i'\hat{\underline{\beta}})$. This is a desirable statistical property in modeling vehicle accidents.[67] Note that most of the suggested conventional multiple linear regression models for establishing geometric design-vehicle accident relationships do not have such a property.[67]

Equation (14) is a set of simultaneous nonlinear equations of $\underline{\beta}$, which needs to be solved numerically. The MLE can be obtained more efficiently by directly maximizing the log-likelihood function in eq (13) through some iterative nonlinear optimization methods, such as the Davidon-Fletcher-Powell algorithm which is employed by this study.[67] Alternative methods include the iterative procedures, based on the quasi-likelihood function of the Poisson model, as that described in McCullagh and Nelder or the generalized least squares procedures introduced in Carroll and Ruppert.[58,57] The maximum likelihood estimates are known to be asymptotically efficient and unbiased (provided, of course, the underlying Poisson distribution and functional specification of eq (10) are correct).

Covariance of the Estimated Parameters

The asymptotic covariance and t-statistics of the estimated parameters can be determined using the second derivative of the loglikelihood function (i.e., the Fisher's information matrix) as follows. The second derivative, or the Hessian matrix, of the loglikelihood function can be derived as:

$$h_{jq} = \frac{\partial^2 L(\underline{\beta})}{\partial \beta_j \partial \beta_q} = -\sum_i (v_i e^{\underline{x}_i \underline{\beta}}) x_{ij} x_{iq} \qquad j=1,2,..,k, \quad q=1,2,...,k \tag{15}$$

which is a function of unknown regression parameter $\underline{\beta}$, and does not involve dependent variable y_i. Provided the Poisson assumption is adequate (with no overdispersion) and the sample size is reasonably large, the asymptotic covariance matrix of the MLE can be obtained as:

$$cov(\hat{\underline{\beta}}) = \left[I(\hat{\underline{\beta}})\right]^{-1} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1k} \\ s_{21} & s_{22} & \cdots & s_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ s_{k1} & s_{k2} & \cdots & s_{kk} \end{bmatrix} \tag{16}$$

where

$$I(\hat{\underline{\beta}}) = -E_y\left\{\left.\frac{\partial^2 L(\underline{\beta})}{\partial \underline{\beta}\, \partial \underline{\beta}'}\right|_{\underline{\beta}=\hat{\underline{\beta}}}\right\} = -\left.\frac{\partial^2 L(\underline{\beta})}{\partial \underline{\beta}\, \partial \underline{\beta}'}\right|_{\underline{\beta}=\hat{\underline{\beta}}} = -\begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{21} & h_{22} & \cdots & h_{2k} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ h_{k1} & h_{k2} & \cdots & h_{kk} \end{bmatrix}_{\underline{\beta}=\hat{\underline{\beta}}} \tag{17}$$

is the Fisher information matrix evaluated at the MLE $\hat{\underline{\beta}}$.[69] The asymptotic t-statistic for each estimated regression parameter $\hat{\beta}_j$ is computed as $\hat{\beta}_j/(s_{jj})^{1/2}$, and its significance level can be assessed using a *t* distribution table with *n-k* degrees of freedom or simply using a normal probability table because of the large *n*. The asymptotic correlation matrix of the estimated regression parameters can be constructed as $\hat{\rho}_{ij}=s_{ij}/(s_{ii}s_{jj})^{1/2}$, for $i=1,2,...,k$, and $j=1,2,..,k$. (Note that $\hat{\rho}_{ii}=1$ for $i=1,2,...,k$.)

Model Selection Criterion

In determining whether the relationship between a specific explanatory variable x_{ij} and truck accident involvement is well-estimated in eq (10), we first checked to see if the estimated parameter $\hat{\beta}_j$ of the explanatory variable had the expected sign, and then we examined whether its t-statistic was greater than 1.96 (or 1.645, or 1.28 for a lower α level). In addition, we used the model selection criterion of Akaike (*AIC*), computed as:

$$AIC(p) = -2L(\hat{\underline{\beta}}) + 2p \tag{18}$$

where $L(\hat{\underline{\beta}})$ is log-likelihood function evaluated at the MLE, and p is the total number of free unknown parameters in the likelihood function, which is equal to k if all k explanatory variables are included in eq (10). Models with smaller *AIC* values are preferred. Bozdogan is an excellent reference on the theory and application of *AIC* criterion.[70]

Goodness-of-Fit Test and Overdispersion

To help assess the overall goodness-of-fit of the proposed model, we considered Pearson's chi-square statistic, X^2, (see, e.g., Agresti).[71] The basic idea of the statistic is to compare the observed frequency, y_i, with the expected frequency, $\hat{\mu}_i$, based on the model. Specifically, the Pearson's X^2 statistic is computed as $\Sigma_i(y_i-\hat{\mu}_i)^2/\hat{\mu}_i$. For large samples, X^2 has approximately a chi-squared distribution with the degrees of freedom (df) equal to the number of sampling units minus the total number of parameters estimated in the model.

In this particular study, where we considered each road section as a sampling unit, the observed and the expected frequencies refer to the observed and the expected number of trucks involved in accidents on each road section, i.e., y_i and $v_i \exp(\underline{x}_i'\hat{\underline{\beta}})$, respectively. The df is therefore n-p, where n is the total number of road sections under consideration and p is the number of regression parameters estimated in the model. Under the hypothesis that the observed and the estimated frequencies come from the same distribution function, X^2 is approximately distributed as a central χ^2 with df of n-p. In cases where the estimated model fails the goodness-of-fit test, e.g., when X^2 is greater than $\chi^2_{0.05}(\mathrm{df}=n\text{-}p)$, it indicates that the Poisson assumption may be inadequate. However, there are many possible reasons for the model to fail the tests, including the overdispersion problem indicated earlier. Another potential difficulty for using the X^2 statistic is that a great percentage of the road sections considered will have no (or very few) accidents. Studies indicate that X^2 statistics may be poorly approximated by a chi-square distribution under this condition.[71] Since this goodness-of-fit test statistic is still being enthusiastically investigated by many researchers for cases where the expected frequencies are small, we suggest that it be used with some caution.

As discussed earlier, a limitation of using the Poisson regression model is that the variance of the data is restrained to be equal to the mean. In many applications, count data were found to display extra variation or overdispersion relative to a Poisson model.[60] That is, the variance of the data was greater than what the Poisson model indicated. The overdispersion could come from several sources, including omitted variables, uncertainties in exposure and explanatory variables, nonhomogeneous environment, and correlations of dependent variables between sampling units.

Following Wedderburn, to correct for the overdispersion problem for the Poisson regression model, one can assume that the variance of Y_i is $\tau\mu_i$ instead of μ_i as that originally assumed in the Poisson model, where τ is called the overdispersion parameter.[66] Under this assumption, the moment estimator of the overdispersion parameter τ is $\hat{\tau} = X^2/(n\text{-}k)$, where X^2 is the Pearson's chi-square statistic, n is the number of observations (i.e., the number of road sections in our case), and k is the number of regression parameters in the Poisson regression model. A better estimate of the asymptotic covariance matrix is $\hat{\tau} \times cov(\hat{\underline{\beta}})$ and, therefore, a better estimate of the t-statistic for regression parameter $\hat{\beta}_j$ is $\hat{\beta}_j/(\hat{\tau}\, s_{jj})^{1/2}$.[71]

In sum, the final selected model should have: (1) the expected signs in all estimated parameters, (2) low *AIC* value, and (3) high t-statistics for model parameters. For the model to be useful in practice, the signs of the parameters should be given the highest consideration in this study. At the same time, explanatory variables with signs in parameters contrary to expectation should be checked and further investigated. If higher order effects of an explanatory variable or interaction effects of two explanatory variables are found to be statistically significant, they should be considered and checked carefully, by, e.g., predicting accident involvement rates outside of the range of the current data to see if the model is still capable of providing reasonable results.

Accident Involvement Reduction Factor

To illustrate how the proposed Poisson regression model can be used to estimate the expected reduction in truck accident involvements due to some improvements in geometric design elements, let's consider a particular road section i, and let the value of its explanatory variables or covariates before and after the improvement be x^b_{ij} and x^a_{ij}, for $j=1,2,..,k$. To illustrate, let x^b_{im} be the horizontal curvature of road section i before the improvement, say, x^b_{im} = 10 degrees, and x^a_{im} be the curvature degree after an improvement, say, x^a_{im} = 3 degrees. Also, let v^b_i and v^a_i be the amount of truck travel in one year on road section i before and after the improvement.

Based on the Poisson regression model (eq (10)), the expected number of truck accident involvements on road section i before and after the improvements of geometric design elements are, respectively, $v_i^b \exp(\Sigma_j x_{ij}^b \beta_j)$ and $v_i^a \exp(\Sigma_j x_{ij}^a \beta_j)$. The percentage reduction in the expected number of truck accident involvements (or truck accident involvement reduction factor) can be computed as:

$$R_i = \left\{ \frac{v_i^b \exp(\sum_{j=1}^{k} x_{ij}^b \beta_j) - v_i^a \exp(\sum_{j=1}^{k} x_{ij}^a \beta_j)}{v_i^b \exp(\sum_{j=1}^{k} x_{ij}^b \beta_j)} \right\} \times 100 \tag{19}$$
$$= \left\{ 1 - \left(\frac{v_i^a}{v_i^b} \right) \exp\left[\sum_{j=1}^{k} (x_{ij}^a - x_{ij}^b) \beta_j \right] \right\} \times 100$$

If v_i is the same before and after the improvement, i.e., $v_i^b = v_i^a$, then R_i also represents the percentage reduction in truck accident involvement *rate*. By substituting β_j with the MLE $\hat{\beta}_j$ in eq (19) for $j=1,2,...,k$, one obtains an MLE of the reduction in the expected number of truck accident involvements, denoted by $\hat{R}_i$. For a large sample $\hat{\underline{\beta}}$ is approximately normally distributed with mean $\underline{\beta}$ and with covariance matrix $\hat{\tau} \times cov(\hat{\underline{\beta}})$. One can show that the standard deviation (*s.d.*) of $\hat{R}_i$ is approximately as follows:

$$s.d.(\hat{R}_i) \approx \left(\frac{v_i^a}{v_i^b} \right) \times \left\{ \exp\left[\sum_{j=1}^{k} (x_{ij}^a - x_{ij}^b) \hat{\beta}_j + \frac{\hat{\tau}}{2} \sum_{m=1}^{k} \sum_{q=1}^{k} (x_{im}^a - x_{im}^b)(x_{iq}^a - x_{iq}^b) s_{mq} \right] \right\} \times \tag{20}$$
$$\left\{ \exp\left[\hat{\tau} \sum_{m=1}^{k} \sum_{q=1}^{k} (x_{im}^a - x_{im}^b)(x_{iq}^a - x_{iq}^b) s_{mq} \right] - 1 \right\}^{1/2} \times 100$$

where, for the Poisson regression considered above, $s_{mq} = \hat{\rho}_{mq} (s_{mm} s_{qq})^{1/2}$. The derivation of this equation utilizes the property that if z is normally distributed with mean μ and variance σ^2, then the variance of $\exp(z)$ is $\{\exp[\mu + (1/2)\sigma^2]\}^2 \{\exp(\sigma^2) - 1\}$. This equation allows one to assess the uncertainty of the estimated reduction by quoting plus or minus one standard deviation.

RESULTS

The primary data source used in this study was the Highway Safety Information System (HSIS). Specifically, highway geometric and traffic data of three roadway types: rural Interstate, urban Interstate and freeway, and rural two-lane undivided arterial, and the associated accidents involving large trucks in Utah from 1985 to 1989, were used for developing the relationships. A detailed description of the data has been presented in chapter 3.

Limited by the number of truck accident involvements available for analysis in this study, one model for each of the three roadway types in Utah was developed for all accident severity types combined. Furthermore, because the truck exposure data are currently not available by truck types, time of day, nor weather conditions, the developed models were not capable of distinguishing the effect of highway geometric design on the safety performance of trucks between different truck types, daytime and nighttime, nor under different weather conditions. Recognizing the limitation of the existing data, the models developed in this study should be considered preliminary. An important objective of this study was, therefore, to understand the uncertainties of the preliminary models, and to suggest ways to improve the models (see chapters 6 and 7).

Explanatory Variables or Covariates

The explanatory variables considered in this study for individual homogeneous road sections, as well as their definitions and summary statistics, are given in tables 13 through 15 for the three roadway classes. Some of these variables have already been introduced in chapter 3. The following is a list of these variables and, when needed, additional discussions are included:

- x_{i1} = 1, representing a dummy intercept for year 1985.
- $x_{i2}, x_{i3}, x_{i4}, x_{i5}$ = Dummy variables to control for year-to-year changes from 1986 to 1989 in the overall truck accident involvement rate due, e.g., to changes in trend, random fluctuations, posted speed limit, and omitted variables such as weather.
- $x_{i6}, x_{i,15}$ = AADT per lane (in thousands of vehicles), a surrogate variable representing vehicle flow density, and its squared term ($x_{i,15}=x_{i6}\times x_{i6}$).
- $x_{i7}, x_{i8}, x_{i,13}$ = Horizontal curvature, HC, (in degrees/100-ft (30.48-m) arc), length of original curve, LHC, (in mi), and their interactions, HC × LHC, i.e., $x_{i,13}=x_{i7}\times x_{i8}$; LHC is considered only for curves with horizontal curvatures greater than 1 degree per 100-ft (30.48-m) arc. As indicated in chapter 3, this definition was based on an assumption that the length of a mild curve (≤ 1 degree) has no aggravated effect on truck accident involvements. In this study, it is also assumed that there was no additional effect of LHC on truck accidents after LHC reaches 1 mi (1.6 km).
- $x_{i9}, x_{i,10}, x_{i,14}$ = Vertical grade, VG, (in percent), length of original grade, LVG, (in mi), and their interactions, VG × LVG, i.e., $x_{i,14}=x_{i9}\times x_{i,10}$; LVG is considered only for grades with vertical grade greater than 2 percent. Again, this definition was based on an assumption that the length of a mild grade (≤ 2 percent) has no aggravated effect

on truck accident involvements. It is also assumed that there was no additional effect of LVG on truck accidents after LVG reaches 2 mi (3.2 km).

- $x_{i,11}$ = Deviation of paved inside or stabilized outside shoulder width (per direction) from an ideal width of 12 ft (3.66 m).

- $x_{i,12}$ = Percent trucks in the traffic stream. This variable is used to evaluate the effect of car-truck mix. Previous studies suggested that as percent trucks increases, truck accident involvement rate decreases. One possible reason is that, *for a constant vehicle density*, as percent trucks increases, the frequency of lane changing and overtaking movements by cars decreases. Also, previous records showed that more trucks are involved in truck-car multivehicle accidents than in truck-truck multivehicle accidents.

- $x_{i,16}$ = A dummy variable to distinguish the overall truck accident involvement rate between urban freeways and urban Interstate highways.

- $x_{i,17}$ = Number of lanes (NL) for urban Interstate and freeway. In this study, NL is between 4 and 8. For road sections with the same traffic density, the ones that have more traffic lanes are expected to have higher truck accident involvement rates than those sections that have less number of traffic lanes. The reason is that, for a constant traffic density, as NL increases the frequency of lane changing and overtaking movements increases.

All of the road sections considered had 12-ft (3.66-m) lane width and almost all road sections on rural Interstate and urban Interstate/freeway had paved outside shoulder widths of 10 ft (3.05 m) per direction, the effects of these variables could not be determined in this study.

Model Estimation and Selection

Several models were tested using different geometric design variables associated with horizontal curvature and vertical grade. The second order or the square term of explanatory variables were also tested when an explanatory variable was considered to have varied over a wide range, e.g., AADT per lane in urban Interstate and freeway and horizontal curvature in rural two-lane undivided arterial. The estimated regression parameters for some of the tested models and the associated standard deviations and asymptotic t-statistics are presented as Models 1 through 3 in tables 16 through 18. The square root of the estimated overdispersion parameter ($\hat{\tau}^{1/2}$), loglikelihood function evaluated at the estimated parameters, L($\hat{\beta}$), and the Akaike Information Criterion (*AIC*) value for each model are also given in these tables. Note that estimated models with high loglikelihood function and low *AIC* values are preferred. Furthermore, the expected total number of trucks involved in accidents across road sections are compared with the observed total. Note that the asymptotic standard deviations and t-statistics presented in these tables are computed using eq (16) without the adjustment of overdispersion.

All models examined failed to pass the chi-square test at a 5-percent α level, especially the one for urban Interstate and freeway. The Wedderburn's overdispersion parameter was therefore used to assess the extent of overdispersion and to adjust the asymptotic standard deviations and t-statistics derived from the MLE. Now, recall that even if the Poisson model is overdispersed the estimates of model parameters, $\underline{\hat{\beta}}$, and therefore the estimated truck accident involvement rate and accident involvements, are still "asymptotically correct." However, as a result of overdispersion, the use of Poisson distribution to predict truck accident involvement probability will be less accurate.

Model 3 in each of these tables has the expected sign in estimated regression parameters and, in most cases, has the lowest AIC value, and is considered the best model among the models tested. In order to examine the effect of short road sections on the estimation of model parameters, we removed road sections with section lengths less than or equal to 0.05 mi (0.08 km). The remaining road sections were then used to recalibrate Model 3. The results are presented as Model 4 in these tables. The comparison of the estimated parameters of Model 3 and Model 4 indicated not only that the conclusions reached regarding the significance level of the relationships between truck accidents and the examined traffic and highway geometric variables were consistent, but also that the estimated parameter values were very close. This suggests that the Poisson regression models are not sensitive to the length of road sections.

Major findings based on the estimated models for each roadway class are presented as follows:

Rural Interstate

1. All of the estimated parameters for the traffic and geometric variables (β_6 through β_{14}) are consistent among different models and have expected algebraic signs.
2. The overdispersion parameters, $\hat{\tau}^{1/2}$, for the four models are 1.25, 1.25, 1.25, and 1.15, respectively, which are not far from unity. This suggests that the truck accident data were overdispersed only moderately over what the estimated Poisson models have implied. On the presumption that truck accident events are Poisson distributed under a perfectly homogeneous highway environment, it appears that not much improvement can be made by bringing more explanatory variables into these models.

3. Based on Model 3, truck accident involvement rates for different combinations of AADT per lane, horizontal curvature, length of original curve, vertical grade, length of original grade, paved inside shoulder width, and percent trucks are illustrated in figure 2. These rates are computed using the average estimated parameters for 1987, 1988, and 1989 dummy variables as intercept. Specifically, they are computed as: $\hat{\lambda}_i = \exp[\hat{\beta}_1 + (\hat{\beta}_3+\hat{\beta}_4+\hat{\beta}_5)/3 + x_{i,6}\hat{\beta}_6 + x_{i,7}\hat{\beta}_7 + x_{i,13}\hat{\beta}_{13} + x_{i,9}\hat{\beta}_9 + x_{i,14}\hat{\beta}_{14} + x_{i,11}\hat{\beta}_{11} + x_{i,12}\hat{\beta}_{12}] = \exp[-0.626471 + 0.0244x_{i,6} + 0.088861x_{i,7} + 0.234209x_{i,13} + 0.077815x_{i,9} + 0.033973x_{i,14} + 0.085763x_{i,11} - 0.025233x_{i,12}]$.

4. For the ranges of covariates indicated in table 13, the final model above suggests the following relationships between traffic/geometric design variables and truck accident involvement rates (using eqs (19) and (20)).

   (1) As AADT per lane increases by 1,000 vehicles per lane, truck accident involvement rate increases by about 2.5 percent.

   (2) As horizontal curvature increases, truck accident involvement rate increases. However, the increase depends on the length of curve. For example, for a curve with 0.1 mi (0.16 km) in length and with curvature greater than 1 degree per 100-ft (30.48-m) arc, as horizontal curvature increases by 1 degree, truck accident involvement rate increases by about 11.9 percent.

   (3) As vertical grade increases, truck accident involvement rate increases. The increase, however, depends on the length of grade. For example, for a grade with 0.5 mi (0.8 km) in length and with vertical grade greater than 2 percent, as vertical grade increases by 1 percent, truck accident involvement rate increases by about 9.9 percent.

   (4) As the length of curve increases, truck accident involvement rate increases. The increase, however, depends on the curvature degree. For example, for a 3-degree curve, as the length of curve increases by 0.1 mi (0.16 km), truck accident involvement rate increases by about 7.3 percent.

   (5) As the length of grade increases, truck accident involvement rate increases. The increase depends on the steepness of the vertical grade. For example, for a 3-percent grade, as the length of grade increases by 0.5 mi (0.8 km), truck accident involvement rate increases by about 5.2 percent.

   (6) As paved inside shoulder width per direction increases by 1 ft (0.3048 m) per direction, truck accident involvement rate decreases by about 8.2 percent.

   (7) *For a constant vehicle density*, as percent trucks in the traffic stream increases by 5 percent, truck accident involvement rate decreases by about 11.9 percent.

For illustration purposes, the asymptotic correlation matrix, $\hat{\rho}_{ij}$, $i=1,2,...,k$, $j=1,2,..,k$, for the estimated parameters of Model 3 is presented in table 19.

Urban Interstate and Freeway

Except the parameter for the length of grade, all of the estimated parameters for the traffic and geometric variables are consistent among different models and have expected algebraic signs. Note that the effects of the length of grade and its interaction with vertical grade on truck accident involvement rate were poorly determined and were, therefore, dropped from the final model. The inclusion of the higher order term for AADT per lane (the square of AADT per lane) was found to improve the model in terms of *AIC* value, but the estimated effect was found to be inconsistent with the results from the other two roadway classes. Therefore, it was decided that only the first order term of AADT per lane be included in the model.

The overdispersion parameters, $\hat{\tau}^{1/2}$, for the four models are 1.71, 1.71, 1.71, and 1.44, respectively. This indicates that the truck accident data exhibit much more dispersion than what the estimated Poisson models have implied. We, therefore, conclude that the developed model can most likely be improved in the future by bringing more explanatory variables into the model, e.g., traffic variables broken down by daytime and nighttime, by increasing the accuracy of truck exposure data, and by better defining the traffic density measure so that congestion related accidents can be better predicted.

Because of the large overdispersion exhibited in the estimated Poisson models, we prefer not to use the model for detailed analyses. However, in the next chapter, we will present an alternative type of model which allows the overdispersion to exist in the data, and detailed analyses about the relationships between truck accidents and highway geometric design for urban Interstate and freeway will be given there.

Rural Two-Lane Undivided Arterial

1. Based on the available data, we were unable to establish reasonable relationships between truck accidents and vertical grade related variables for rural two-lane undivided arterial. For example, the estimated parameter of vertical grade has a negative parameter value, which is contrary to what one would expect. Overall, the estimated models for rural two-lane undivided arterial were not as well determined as those for rural Interstate examined in this study. The main reason is that during the period examined rural two-lane undivided arterial has substantially less number of truck accident involvements and truck

travel available for analysis than those available for rural Interstate. Although the estimated parameter for the square of horizontal curvature was not reported in table 18, it was tested and found to be insignificant at a 10-percent α level.

2. The overdispersion parameters, $\hat{\tau}^{1/2}$, for the four models are 1.36, 1.37, 1.36, and 1.27, respectively. This suggests that the truck accident data exhibit moderate overdispersion over the estimated Poisson models. It is, therefore, expected that the developed model can probably be improved in the future by collecting more truck miles, by bringing more explanatory variables into the model, and by increasing the accuracy of truck exposure data.

3. The final selected model includes the following variables: AADT per lane, horizontal curvature, length of original curve, stabilized outside shoulder width, and percent trucks. Truck accident involvement rate can be computed as: $\hat{\lambda}_i = \exp[\hat{\beta}_1 + (\hat{\beta}_3+\hat{\beta}_4+\hat{\beta}_5)/3 + x_{i,6}\hat{\beta}_6 + x_{i,7}\hat{\beta}_7 + x_{i,13}\hat{\beta}_{13} + x_{i,11}\hat{\beta}_{11} + x_{i,12}\hat{\beta}_{12}] = \exp[0.0817646 + 0.102226x_{i,6} + 0.094931x_{i,7} + 0.042564x_{i,13} + 0.034061x_{i,11} - 0.026276x_{i,12}]$.

4. For the ranges of covariates indicated in table 15, the above model suggests the following relationships between traffic/geometric design variables and truck accident involvement rates.

   (1) As AADT per lane increases by 1,000 vehicles per direction, truck accident involvement rate increases by 10.8 percent.

   (2) As horizontal curvature increases, truck accident involvement rate increases. However, the increase depends on the length of curve. For example, for a curve with 0.1 mi (0.16 km) in length and with curvature greater than 1 degree per 100-ft (30.48-m) arc, as horizontal curvature increases by 1 degree, truck accident involvement rate increases by about 10.4 percent.

   (3) As the length of curve increases, truck accident involvement rate increases. The increase, however, depends on the curvature degree. For example, for a 5-degree curve, as the length of curve increases by 0.1 mi (0.16 km), truck accident involvement rate increases by about 2.2 percent.

   (4) As stabilized outside shoulder width increases by 1 ft (0.3048 m) per direction, truck accident involvement rate decreases by about 3.3 percent.

   (5) *For a constant vehicle density*, as percent trucks in the traffic stream increases by 1 percent, truck accident involvement rate decreases by about 2.6 percent.

Table 13. Variable definitions and summary statistics of the 8,263 rural Interstate road sections.

Variable	Notation & Definition (for section i)	Min	Max	Mean	% Zero
Number of Trucks Involved in Accidents	y_i	0	8	0.20	86
Section Length (in mi)	ℓ_i	0.01	7.77	0.45	0
Truck Miles or Truck Exposure (in 10^6 truck-miles)	$v_i = [365 \times AADT_i \times (T\%_i/100) \times \ell_i]/10^6$, where $T\%_i$ is percent trucks (366 for leap years).	8×10^{-4}	5.03	0.25	0
Dummy Intercept	$x_{i1} = 1$				
Dummy Variable for Year 1986, representing year-to-year changes due to random fluctuations, annual trend, and omitted variables such as weather.	$x_{i2} = 1$, if the road section is in year 1986 $= 0$, otherwise				
Dummy Variable for Year 1987 (See above explanation)	$x_{i3} = 1$, if the section is in 1987 $= 0$, otherwise				
Dummy Variable for Year 1988 (See above explanation)	$x_{i4} = 1$, if the section is in 1988 $= 0$, otherwise				
Dummy Variable for Year 1989 (See above explanation)	$x_{i5} = 1$, if the section is in 1989 $= 0$, otherwise				
AADT per Lane (in 1000's of vehicles), a surrogate variable to indicate traffic conditions or traffic density.	$x_{i6} = (AADT_i/\text{number of lanes}_i)/1000$	0.35	12.04	1.80	0
Horizontal Curvature, HC, (in degrees per 100-ft arc)	x_{i7}	0	12.00	1.00	67
Length of Original Horizontal Curve, LHC, (in mi) from which this curve was subdivided for creating homogeneous sections; only for HC > 1 and LHC ≤ 1.	x_{i8} = LHC, if $x_{i7}>1$ and LHC ≤ 1 mi. = 1.0, if $x_{i7}>1$ and LHC > 1 mi. = 0, if $x_{i7} \leq 1$	0	0.96	0.05	81
Vertical Grade, VG, (in percent)	x_{i9}	0	8.00	2.14	20
Length of Original Vertical Grade, LVG, (in mi) from which this section was subdivided for creating homogeneous sections; only for sections with VG > 2 and LVG ≤ 2.	$x_{i,10}$ = LVG, if $x_{i9}>2$ and LVG ≤ 2 mi. = 2.0, if $x_{i9}>2$ and LVG > 2 mi. = 0, if $x_{i9} \leq 2$	0	2.00	0.21	74
Deviation of Paved Inside Shoulder Width (per direction) from an "ideal" width of 12 ft (3.66 m).	$x_{i,11}$ = max{0, 12 - paved inside shoulder width}	4.00	12.00	8.16	0
Percent Trucks in the traffic stream (e.g., 15)	$x_{i,12}$	7.00	57.00	24.13	0
HC × LHC	$x_{i,13} = x_{i7} \times x_{i8}$	0	2.88	0.18	81
VG × LVG	$x_{i,14} = x_{i9} \times x_{i,10}$	0	13.37	0.97	74

(1) All of the sections are 12 ft (3.66 m) in lane width. (2) About 89% of the sections have 4 lanes; and all sections have paved outside shoulder width of 10 ft (3.05 m). (3) Total number of trucks involved in accidents = 1,643; total highway lane-mi = 14,731; and total truck travel = $2{,}030 \times 10^6$ truck mi ($3\,248 \times 10^6$ truck km). (4) 1 mi = 1.61 km; 1 ft = 0.3048 m.

Table 14. Variable definitions and summary statistics of the 2,810 urban Interstate and freeway road sections.

Variable	Notation & Definition (for section i)	Min	Max	Mean	% Zero
Number of Trucks Involved in Accidents	y_i	0	19	0.68	67
Section Length (in mi)	ℓ_i	0.01	3.02	0.26	0
Truck Miles or Truck Exposure (in 10^6 truck-miles)	$v_i = [365 \times AADT_i \times (T\%_i/100) \times \ell_i]/10^6$, where $T\%_i$ is percent trucks (366 for leap years).	1.5×10^{-3}	6.03	0.37	0
Dummy Intercept	$x_{i1} = 1$				
Dummy Variable for Year 1986, representing year-to-year changes due to random fluctuations, annual trend, and omitted variables such as weather.	$x_{i2} = 1$, if the road section is in year 1986 $= 0$, otherwise				
Dummy Variable for Year 1987 (See above explanation)	$x_{i3} = 1$, if the section is in 1987 $= 0$, otherwise				
Dummy Variable for Year 1988 (See above explanation)	$x_{i4} = 1$, if the section is in 1988 $= 0$, otherwise				
Dummy Variable for Year 1989 (See above explanation)	$x_{i5} = 1$, if the section is in 1989 $= 0$, otherwise				
AADT per Lane (in 1000's of vehicles), a surrogate variable to indicate traffic conditions or traffic density.	$x_{i6} = (AADT_i/\text{number of lanes}_i)/1000$	0.62	21.67	7.69	0
Horizontal Curvature, HC, (in degrees per 100-ft arc)	x_{i7}	0	9.00	0.84	67
Length of Original Horizontal Curve, LHC, (in mi) from which this curve was subdivided for creating homogeneous sections; only for HC > 1 and LHC ≤ 1.	x_{i8} = LHC, if $x_{i7}>1$ and LHC ≤ 1 mi. = 1.0, if $x_{i7}>1$ and LHC > 1 mi. = 0, if $x_{i7} \leq 1$	0	1.00	0.09	78
Vertical Grade, VG, (in percent)	x_{i9}	0	12.00	2.07	14
Length of Original Vertical Grade, LVG, (in mi) from which this section was subdivided for creating homogeneous sections; only for sections with VG > 2 and LVG ≤ 2.	$x_{i,10}$ = LVG, if $x_{i9}>2$ and LVG ≤ 2 mi. = 2.0, if $x_{i9}>2$ and LVG > 2 mi. = 0, if $x_{i9} \leq 2$	0	1.25	0.06	86
Deviation of Paved Inside Shoulder Width (per direction) from an "ideal" width of 12 ft (3.66 m).	$x_{i,11}$ = max{0, 12 - paved inside shoulder width}	2.00	12.00	8.01	0
Percent Trucks in the traffic stream (e.g., 5)	$x_{i,12}$	1.00	35.00	11.00	0
HC × LHC	$x_{i,13} = x_{i7} \times x_{i8}$	0	3.00	0.26	78
VG × LVG	$x_{i,14} = x_{i9} \times x_{i,10}$	0	6.00	0.28	86
(AADT per Lane)²	$x_{i,15} = x_{i6} \times x_{i6}$	0.38	469.6	78.77	0
Dummy Variable for Urban Freeways	$x_{i,16}$ = 1, if an urban freeway section (non-Interstate) = 0, otherwise				
Number of Lanes, NL	$x_{i,17}$	4.00	8.00	5.45	0

(1) All of the sections are 12 ft (3.66 m) in lane width. (2) About 95% of sections have the paved outside shoulder width per direction of 10 ft (3.05 m). (3) Total number of trucks involved in accidents = 2,035; total highway lane-mi = 3,889; and total truck travel = $1{,}044 \times 10^6$ truck mi ($1\,670 \times 10^6$ truck km). (4) 1 mi = 1.61 km; 1 ft = 0.3048 m.

Table 15. Variable definitions and summary statistics of the 13,634 rural two-lane undivided arterial road sections.

Variable	Notation & Definition (for section i)	Min	Max	Mean	% Zero
Number of Trucks Involved in Accidents	y_i	0	7	0.06	95
Section Length (in mi)	ℓ_i	0.01	9.41	0.34	0
Truck Miles or Truck Exposure (in 10^6 truck-miles)	$v_i = [365 \times AADT_i \times (T\%_i/100) \times \ell_i]/10^6$, where $T\%_i$ is percent trucks (366 for leap years).	4×10^{-4}	1.95	0.05	0
Dummy Intercept	$x_{i1} = 1$				
Dummy Variable for Year 1986, representing year-to-year changes due to random fluctuations, annual trend, and omitted variables such as weather.	x_{i2} = 1, if the road section is in year 1986 = 0, otherwise				
Dummy Variable for Year 1987 (See above explanation)	x_{i3} = 1, if the section is in 1987 = 0, otherwise				
Dummy Variable for Year 1988 (See above explanation)	x_{i4} = 1, if the section is in 1988 = 0, otherwise				
Dummy Variable for Year 1989 (See above explanation)	x_{i5} = 1, if the section is in 1989 = 0, otherwise				
AADT per Lane (in 1000's of vehicles), a surrogate variable to indicate traffic conditions or traffic density.	$x_{i6} = (AADT_i/\text{number of lanes}_i)/1000$	0.10	8.10	1.26	0
Horizontal Curvature, HC, (in degrees per 100-ft arc)	x_{i7}	0	30.00	2.38	67
Length of Original Horizontal Curve, LHC, (in mi) from which this curve was subdivided for creating homogeneous sections; only for HC > 1 and LHC ≤ 1.	x_{i8} = LHC, if $x_{i7}>1$ and LHC ≤ 1 mi. = 1.0, if $x_{i7}>1$ and LHC > 1 mi. = 0, if $x_{i7}\leq1$	0	0.84	0.05	67
Vertical Grade, VG, (in percent)	x_{i9}	0	11.00	2.70	5
Length of Original Vertical Grade, LVG, (in mi) from which this section was subdivided for creating homogeneous sections; only for sections with VG > 2 and LVG ≤ 2.	$x_{i,10}$ = LVG, if $x_{i9}>2$ and LVG ≤ 2 mi. = 2.0, if $x_{i9}>2$ and LVG > 2 mi. = 0, if $x_{i9}\leq2$	0	2.00	0.19	70
Deviation of Stabilized Outside Shoulder Width (per direction) from an "ideal" width of 12 ft (3.66 m).	$x_{i,11}$ = max{0, 12 - stabilized outside shoulder width}	0	12.00	7.76	<1
Percent Trucks in the traffic stream (e.g., 15)	$x_{i,12}$	1.00	56.00	17.62	0
HC × LHC	$x_{i,13} = x_{i7} \times x_{i8}$	0	5.94	0.30	67
VG × LVG	$x_{i,14} = x_{i9} \times x_{i,10}$	0	15.18	0.91	70

(1) All of the sections are 12 ft (3.66 m) in lane width. (2) Total number of trucks involved in accidents = 789; total highway lane-mi = 9,211; and total truck travel = 694×10^6 truck mi (1 110×10^6 truck km). (3) 1 mi = 1.61 km; 1 ft = 0.3048 m.

Table 16. Estimated parameters of the tested Poisson regression models and associated statistics: rural Interstate.

	Model 1	Model 2	Model 3	Model 4
Section length & number of sections	≥0.01 mi 8,263	≥0.01 mi 8,263	≥0.01 mi 8,263	>0.05 mi 7,004
β_1 Dummy intercept	-0.472330 (±0.287;-1.65)	-0.472494 (±0.287;-1.65)	-0.431762 (±0.288;-1.50)	-0.526103 (±0.290;-1.81)
β_2 Dummy variable for 1986	-0.182576 (±0.086;-2.12)	-0.185384 (±0.086;-2.15)	-0.183853 (±0.086;-2.14)	-0.171759 (±0.087;-1.97)
β_3 Dummy variable for 1987	-0.160249 (±0.085;-1.89)	-0.162656 (±0.085;-1.91)	-0.161461 (±0.085;-1.90)	-0.160869 (±0.086;-1.86)
β_4 Dummy variable for 1988	-0.114524 (±0.085;-1.35)	-0.112753 (±0.085;-1.33)	-0.111511 (±0.085;-1.31)	-0.096243 (±0.086;-1.12)
β_5 Dummy variable for 1989	-0.315484 (±0.088;-3.57)	-0.313863 (±0.088;-3.57)	-0.311155 (±0.088;-3.54)	-0.299701 (±0.089;-3.36)
β_6 AADT per lane (10^3)	0.026710 (±0.015;1.73)	0.022138 (±0.015;1.38)	0.024400 (±0.015;1.59)	0.025220 (±0.015;1.63)
β_7 Horizontal curvature	0.147259 (±0.022;6.85)	0.089178 (±0.028;3.15)	0.088861 (±0.028;3.14)	0.096170 (±0.029;3.27)
β_8 Length of original curve	0.004148 (±0.232;0.02)	-----	-----	-----
β_{13} (Horizontal curvature) × (Length of original curve)	-----	0.232377 (±0.084;2.76)	0.234209 (±0.084;2.78)	0.221877 (±0.087;2.56)
β_9 Vertical grade	0.083423 (±0.027;3.06)	0.084194 (±0.027;3.09)	0.077815 (±0.028;2.81)	0.078218 (±0.028;2.78)
β_{10} Length of original grade	0.165342 (±0.078;2.11)	0.156212 (±0.078;1.99)	-----	-----
β_{14} (Vertical grade) × (Length of original grade)	-----	-----	0.033973 (±0.015;2.26)	0.031085 (±0.015;2.03)
β_{11} Deviation of paved inside shoulder width per direction from 12 ft	0.088652 (±0.036;2.46)	0.091478 (±0.036;2.54)	0.085763 (±0.036;2.37)	0.094814 (±0.036;2.60)
β_{12} Percent trucks (e.g., 15)	-0.025260 (±0.004;-5.91)	-0.025738 (±0.004;-6.01)	-0.025233 (±0.004;-5.88)	-0.025308 (±0.004;-5.82)
$\hat{\tau}^{1/2}$	1.25	1.25	1.25	1.15
$L(\hat{\beta})$	-3775.3	-3771.7	-3771.0	
AIC Value	7574.5	7567.3	7566.0	
Expected vs. Observed Total Truck Accident Involvements	1,642.3 1,643.0	1,644.2 1,643.0	1,644.3 1,643.0	1,604.5 1,603.0

Notes: (1) Values in parentheses are (unadjusted) standard deviation and t-statistics of the parameters above.
(2) ----- Not included in the model.
(3) Model 3 is the final model selected for further analyses.
(4) 1 mi = 1.61 km, 1 ft = 0.3048 m.

Table 17. Estimated parameters of the tested Poisson regression models and associated statistics: urban Interstate and freeway.

	Model 1	Model 2	Model 3	Model 4
Section length & number of sections	≥0.01 mi 2,810	≥0.01 mi 2,810	≥0.01 mi 2,810	>0.05 mi 2,271
β_1 Dummy intercept	-0.933473 (±0.389;-2.40)	-0.947077 (±0.390;-2.42)	-0.939656 (±0.389;-2.42)	-1.09700 (±0.399;-2.75)
β_2 Dummy variable for 1986	-0.385411 (±0.073;-5.28)	-0.386047 (±0.073;-5.28)	-0.385215 (±0.073;-5.27)	-0.384856 (±0.076;-5.06)
β_3 Dummy variable for 1987	-0.582285 (±0.076;-7.71)	-0.583123 (±0.076;-7.72)	-0.582372 (±0.076;-7.71)	-0.579572 (±0.079;-7.37)
β_4 Dummy variable for 1988	-0.293108 (±0.069;-4.25)	-0.292705 (±0.069;-4.24)	-0.292152 (±0.069;-4.24)	-0.266879 (±0.069;-3.47)
β_5 Dummy variable for 1989	-0.273871 (±0.067;-4.10)	-0.273458 (±0.067;-4.09)	-0.273012 (±0.067;-4.09)	-0.239820 (±0.069;-3.47)
β_{16} Dummy variable for Freeways	1.40569 (±0.126;11.19)	1.41983 (±0.134;10.56)	1.40603 (±0.123;11.43)	1.43477 (±0.127;11.27)
β_6 AADT per lane (10^3)	0.045849 (±0.006;8.10)	0.045938 (±0.006;8.11)	0.046010 (±0.006;8.13)	0.046476 (±0.006;7.87)
β_{17} Number of lanes	0.125364 (±0.031;4.00)	0.124321 (±0.031;3.97)	0.124950 (±0.031;4.00)	0.122212 (±0.033;3.76)
β_7 Horizontal curvature	0.034455 (±0.027;1.29)	0.017120 (±0.036;0.47)	0.016375 (±0.036;0.45)	0.006064 (±0.038;0.16)
β_8 Length of the original curve	0.278753 (±0.205;1.36)	------	------	------
β_{13} Horizontal curvature × Length of original curve	------	0.127572 (±0.089;1.43)	0.128738 (±0.089;1.45)	0.138446 (±0.093;1.49)
β_9 Vertical grade	0.099904 (±0.036;2.75)	0.109044 (±0.045;2.42)	0.101143 (±0.033;3.05)	0.096215 (±0.035;2.75)
β_{10} Length of the original grade	-0.006499 (±0.159;-0.04)	------	------	------
β_{14} Vertical grade × Length of original grade	------	-0.014603 (±0.055;-0.26)	------	------
β_{11} Deviation of paved inside shoulder width per direction from 12 ft	0.153878 (±0.041;3.77)	0.153576 (±0.041;3.75)	0.153900 (±0.041;3.77)	0.161322 (±0.042;3.87)
β_{12} Percent Trucks (e.g., 5)	-0.094672 (±0.008;-11.72)	-0.093867 (±0.008;-11.67)	-0.093899 (±0.008;-11.67)	-0.087622 (±0.008;-10.53)
$\hat{\tau}^{1/2}$	1.71	1.71	1.71	1.44
$L(\hat{\underline{\beta}})$	-2742.0	-2741.8	-2741.9	
AIC Value	5512.0	5511.7	5509.7	
Expected vs. Observed Total Truck Accident Involvements	1,904.6 1,904.0	1,903.7 1,904.0	1,903.9 1,904.0	1,773.9 1,774.0

Notes: (1) Values in parentheses are (unadjusted) standard deviation and t-statistics of the parameters above.
(2) ----- Not included in the model.
(3) Model 3 is the final model selected for further analyses.
(4) 1 mi = 1.61 km, 1 ft = 0.3048 m.

Table 18. Estimated parameters of the tested Poisson regression models and associated statistics: rural two-lane undivided arterial.

	Model 1	Model 2	Model 3	Model 4
Section length & number of sections	≥0.01 mi 13,634	≥0.01 mi 13,634	≥0.01 mi 13,634	>0.05 mi 10,371
β_1 Dummy intercept	0.622121 (±0.222;2.80)	0.620408 (±0.224;2.78)	0.434863 (±0.204;2.13)	0.260334 (±0.209;1.25)
β_2 Dummy variable for 1986	-0.182812 (±0.109;-1.68)	-0.178086 (±0.109;-1.64)	-0.181743 (±0.109;-1.67)	-0.196441 (±0.111;-1.76)
β_3 Dummy variable for 1987	-0.463729 (±0.117;-3.95)	-0.458520 (±0.117;-3.91)	-0.461769 (±0.117;-3.94)	-0.446411 (±0.119;-3.74)
β_4 Dummy variable for 1988	-0.352223 (±0.108;-3.25)	-0.347608 (±0.109;-3.20)	-0.344109 (±0.108;-3.17)	-0.333153 (±0.111;-3.01)
β_5 Dummy variable for 1989	-0.260245 (±0.106;-2.46)	-0.256203 (±0.106;-2.42)	-0.253417 (±0.106;-2.40)	-0.225891 (±0.107;-2.10)
β_6 AADT per lane (10^3)	0.093753 (±0.037;2.51)	0.098779 (±0.037;2.65)	0.102226 (±0.037;2.76)	0.123219 (±0.038;3.25)
β_7 Horizontal curvature	0.101750 (±0.010;10.21)	0.095118 (±0.015;6.22)	0.094931 (±0.015;6.16)	0.092632 (±0.017;5.41)
β_8 Length of original curve	0.094085 (±0.037;2.51)	-----	-----	-----
β_{13} (Horizontal curvature) × (Length of original curve)	-----	0.060314 (±0.102;0.59)	0.042564 (±0.100;0.42)	0.080680 (±0.106;0.76)
β_9 Vertical grade	-0.093252 (±0.043;-2.16)	-0.106317 (±0.048;-2.18)	-----	-----
β_{10} Length of original grade	0.149544 (±0.095;1.57)	-----	-----	-----
β_{14} (Vertical grade) × (Length of original grade)	-----	0.040368 (±0.027;1.48)	-----	-----
β_{11} Deviation of stabilized outside shoulder width per direction from 12 ft	0.037167 (±0.015;2.54)	0.038251 (±0.015;2.61)	0.034061 (±0.015;2.34)	0.039412 (±0.015;2.66)
β_{12} Percent trucks (e.g., 15)	-0.026684 (±0.005;-5.49)	-0.026198 (±0.005;-5.38)	-0.026276 (±0.005;-5.38)	-0.022252 (±0.005;-4.46)
$\hat{\tau}^{1/2}$	1.36	1.37	1.36	1.27
$L(\hat{\beta})$	-2543.4	-2543.3	-2545.8	
AIC Value	5110.8	5110.6	5111.6	
Expected vs. Observed Total Truck Accident Involvements	789.2 789.0	789.7 789.0	788.9 789.0	759.7 760.0

Notes: (1) Values in parentheses are (unadjusted) standard deviation and t-statistics of the parameters above.
(2) ----- Not included in the model.
(3) Model 3 is the final model selected for further analyses.
(4) 1 mi = 1.61 km, 1 ft = 0.3048 m.

Table 19. Asymptotic correlation matrix, $[\hat{\rho}_{ij}]$, of the estimated regression parameters, $\hat{\boldsymbol{\beta}}$: rural Interstate (Model 3 in table 16).

	1	2	3	4	5	6	7	13	9	14	11	12
1	1.000	-0.095	-0.087	-0 083	-0.054	-0.193	-0.059	0.034	-0.080	0.099	-0.905	-0.058
2	-0.095	1.000	0.564	0.586	0.578	-0.169	-0.026	0.017	-0.001	-0.029	0.059	-0.305
3	-0.087	0.564	1.000	0.601	0.594	-0.194	-0.029	0.018	0.004	-0.034	0.058	-0.324
4	-0.083	0.586	0.601	1.000	0.628	-0.256	-0.037	0.025	0.012	-0.048	0.080	-0.401
5	-0.054	0.578	0.594	0.628	1.000	-0.314	-0.039	0.029	0.013	-0.053	0.070	-0.425
6	-0.193	-0.169	-0.194	-0.256	-0.314	1.000	0.047	-0.076	-0.070	0.127	-0.076	0.590
7	-0.059	-0.026	-0.029	-0.037	-0.039	0.047	1.000	-0.792	-0.050	0.024	0.009	0.106
13	0.034	0.017	0.018	0.025	0.029	-0.076	-0.792	1.000	-0.007	-0.020	-0.003	-0.069
9	-0.080	-0.001	0.004	0.012	0.013	-0.070	-0.050	-0.007	1.000	-0.783	-0.068	0.001
14	0.099	-0.029	-0.034	-0.048	-0.053	0.127	0.024	-0.020	-0.783	1.000	-0.041	0.113
11	-0.905	0.059	0.058	0.080	0.070	-0.076	0.009	-0.003	-0.068	-0.041	1.000	-0.281
12	-0.058	-0.305	-0.324	-0.401	-0.425	0.590	0.106	-0.069	0.001	0.113	-0.281	1.000

In this figure, ISH, T%, VG, and LVG represent paved inside shoulder width, percent trucks, vertical grade, and length of original grade, respectively. Lines 1 through 10 in each part of the figure show such relationships for different combinations of horizontal curvature (HC) in degrees per 100-ft (30.48-m) arc and length of original curve (LHC) in miles:

Line 1: HC=0;
Line 2: HC=3, LHC=0.1; Line 3: HC=3, LHC=0.5; Line 4: HC=3, LHC=1.0;
Line 5: HC=6, LHC=0.1; Line 6: HC=6, LHC=0.5; Line 7: HC=6, LHC=1.0;
Line 8: HC=9, LHC=0.1; Line 9: HC=9, LHC=0.5; Line 10: HC=9, LHC=1.0.

This figure applies mainly to road sections with 12 ft (3.66 m) lane width and 10 ft (3.05 m) paved outside shoulder width. In each part of the figure, the line numbers from the bottom to the top are: 1, 2, 3, 5, 4, 8, 6, 9, 7, and 10.

(MTM = million truck miles or 1.6 million truck kilometers, 1 ft = 0.3048 m, and 1 mi = 1.61 km)

Figure 2. The relationship between truck accident involvement rate and highway geometric design for rural Interstate highways, based on Model 3 in table 16.

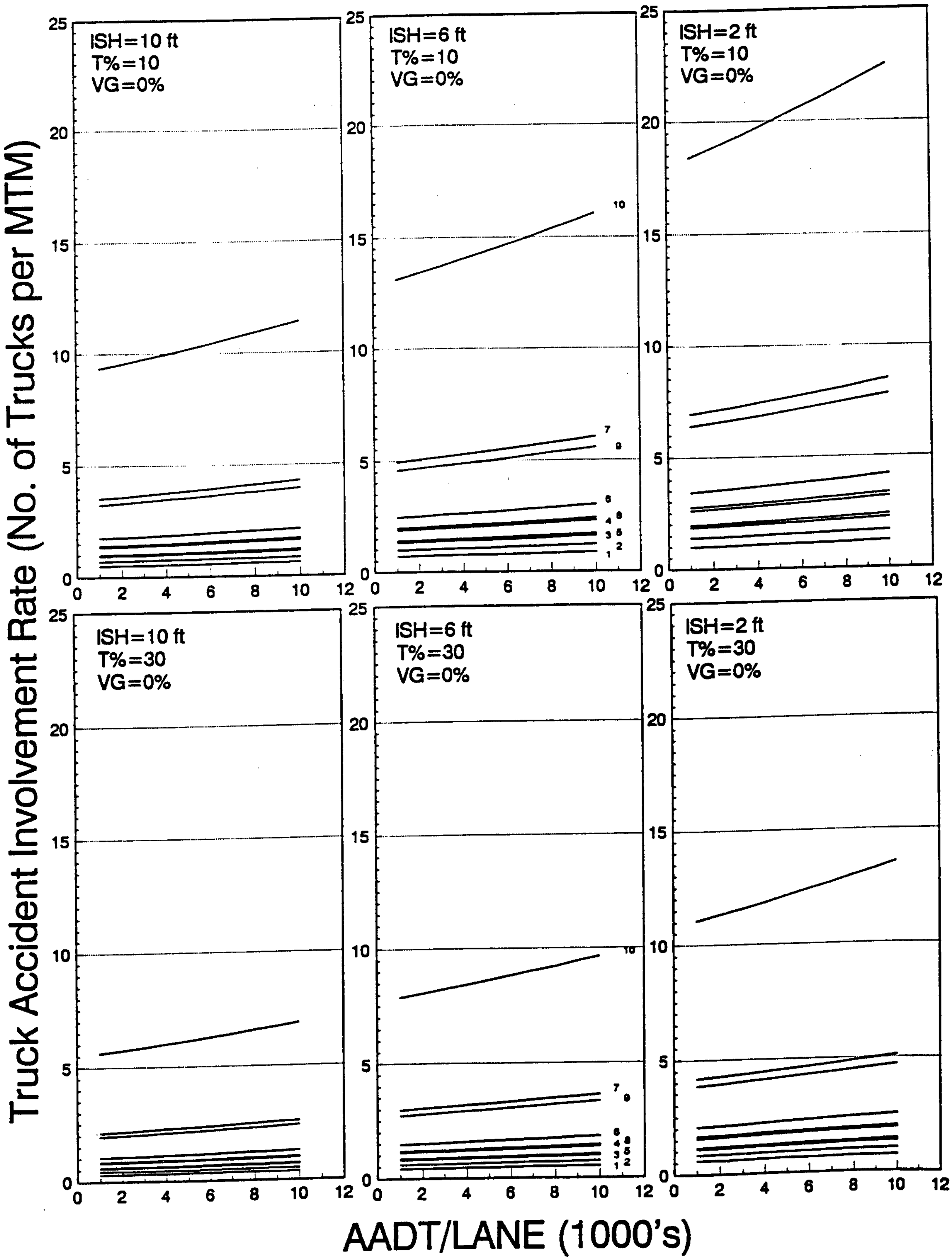

Figure 2. The relationship between truck accident involvement rate and highway geometric design for rural Interstate highways, based on Model 3 in table 16 (continued).

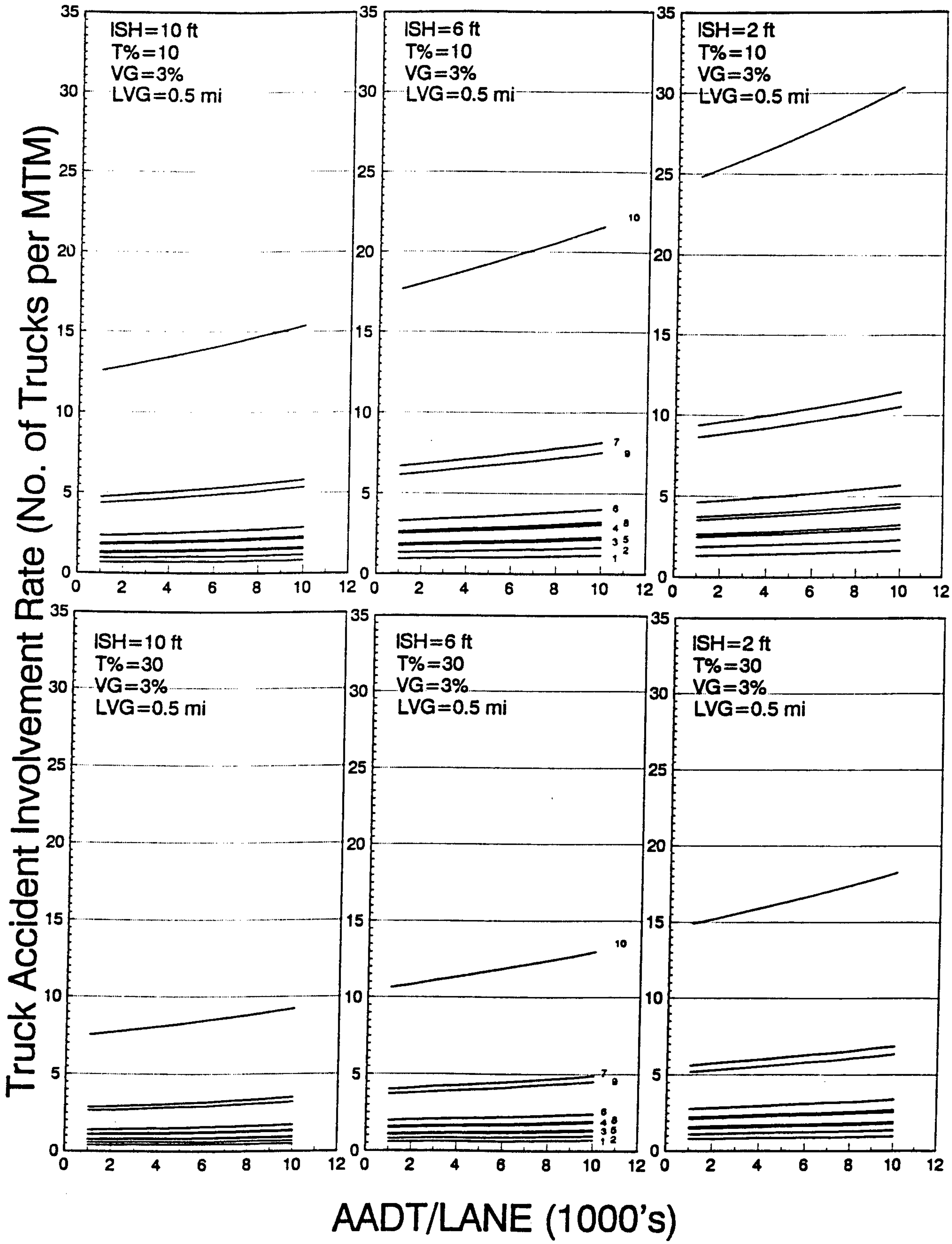

Figure 2. The relationship between truck accident involvement rate and highway geometric design for rural Interstate highways, based on Model 3 in table 16 (continued).

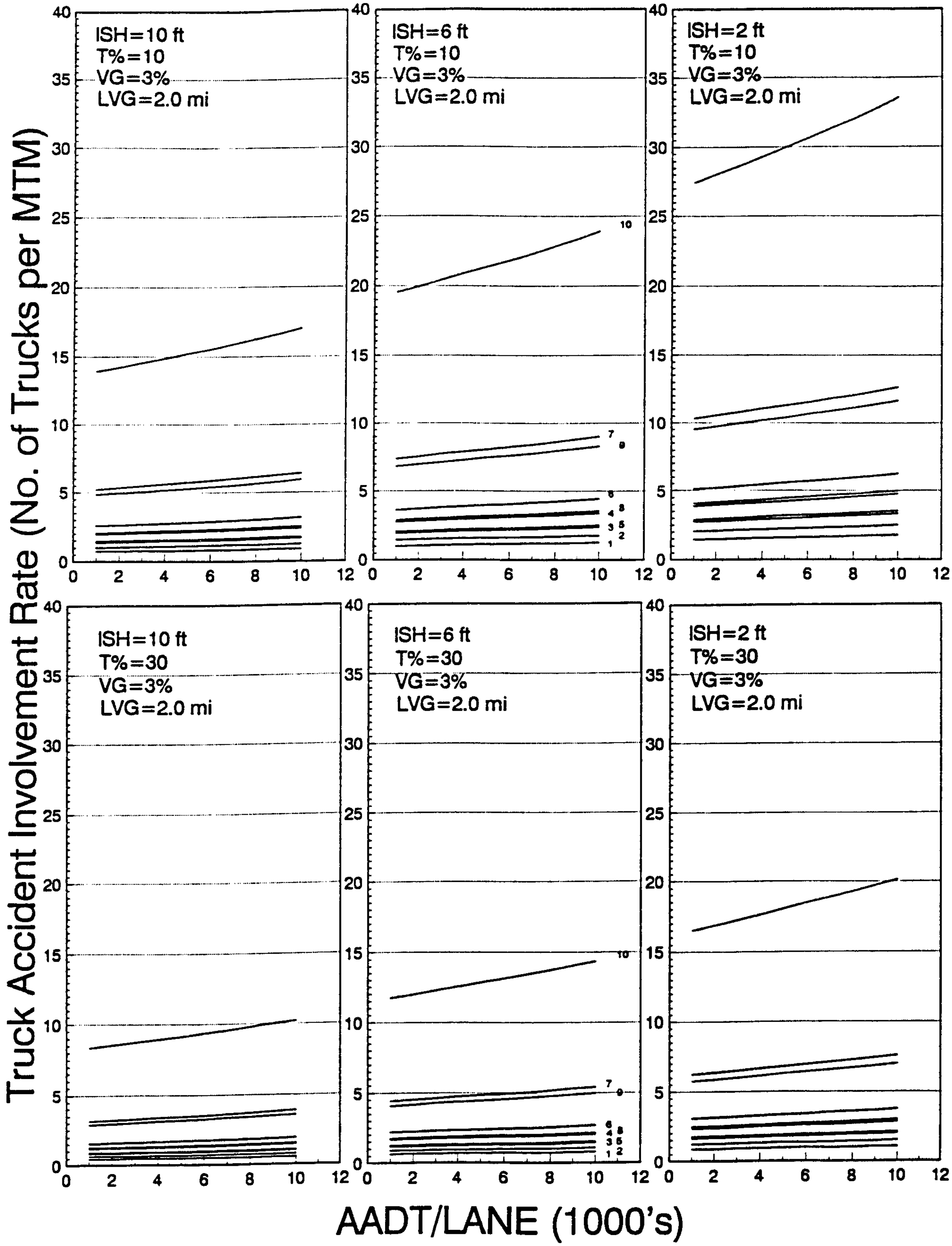

Figure 2. The relationship between truck accident involvement rate and highway geometric design for rural Interstate highways, based on Model 3 in table 16 (continued).

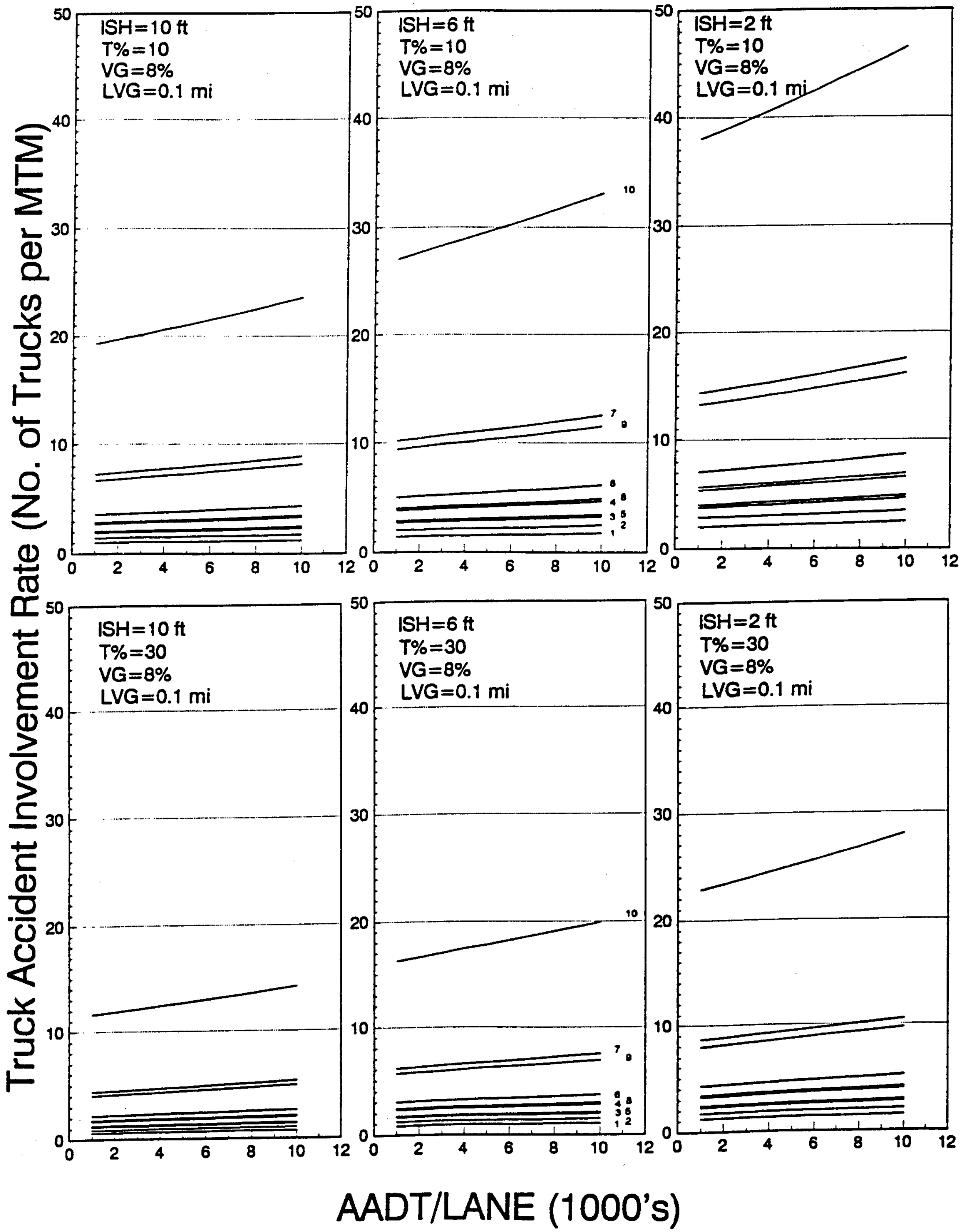

Figure 2. The relationship between truck accident involvement rate and highway geometric design for rural Interstate highways, based on Model 3 in table 16 (continued).

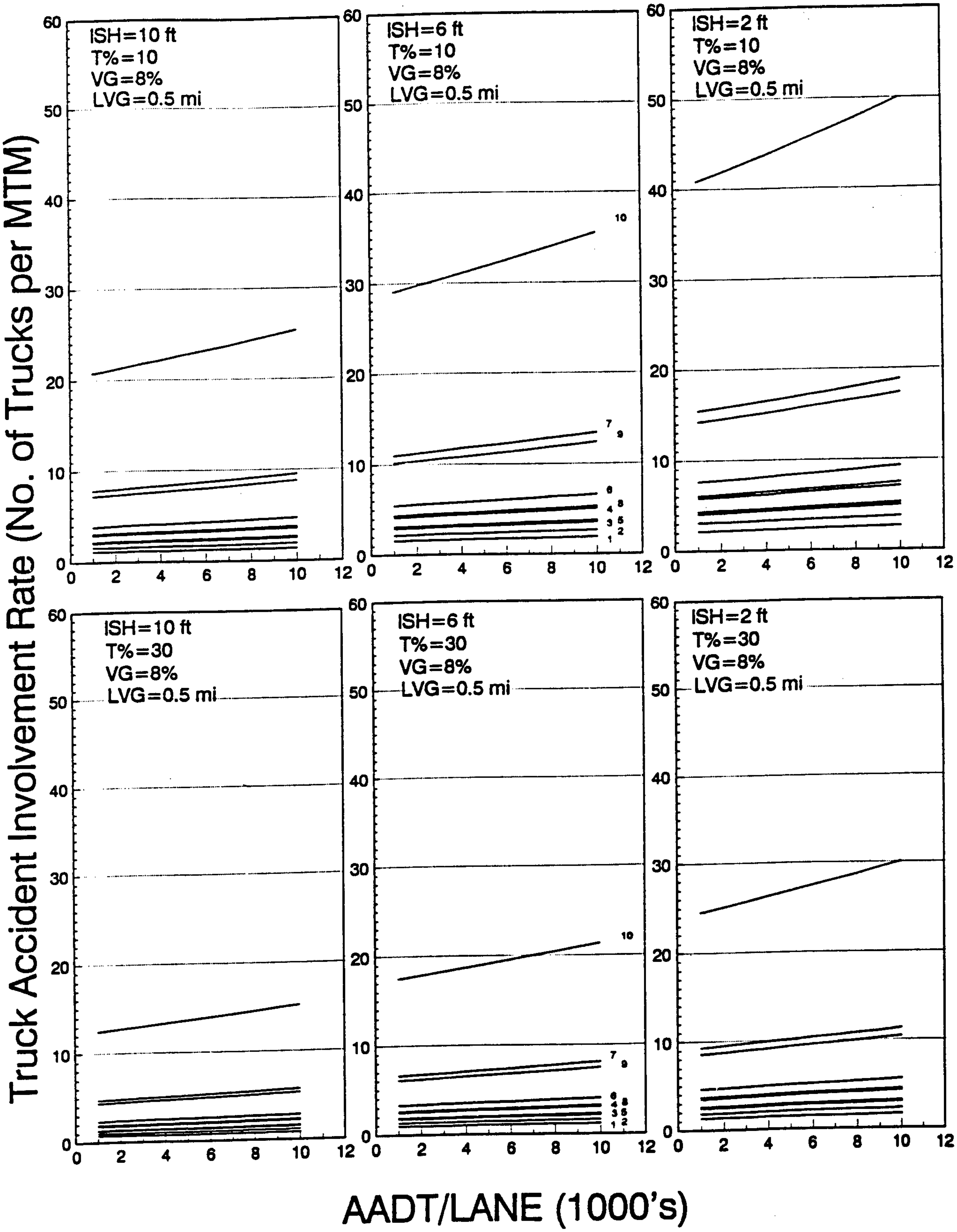

Figure 2. The relationship between truck accident involvement rate and highway geometric design for rural Interstate highways, based on Model 3 in table 16 (continued).

EXAMPLES

Despite the limitations in existing Utah geometric design data, e.g., lack of variation in lane width, some encouraging relationships were developed for horizontal curvature, vertical grade, and shoulder width, using the Poisson regression models. The uncertainties associated with these models are still quite large, especially, for the models for urban Interstate and freeway and rural two-lane undivided arterial. Therefore, we would like to reemphasize here that the models developed in this study should still be considered preliminary. Also, we recommend that these models be used with some caution. This section is intended only to give some examples.

Accident Involvement Rates and Relative Risk

Although this study did not distinguish accidents by truck types, in principal, the overall framework could be applied to any truck type, provided that there are enough accident data and that truck exposure data by truck type are available for individual road sections. Thus, one of the potential applications of the proposed models is to determine the "relative risk" between single-unit and combination trucks when traveling through a given road section. For example, if the estimated rates from the Poisson regression models for single-units and combinations are respectively $\hat{\lambda}_s$ and $\hat{\lambda}_c$, the relative risk between a single-unit truck and a combination truck traveling through a road section with length ℓ can be computed as $[1 - \exp(-\hat{\lambda}_s\ell)]/[1 - \exp(-\hat{\lambda}_c\ell)]$, where $[1 - \exp(-\hat{\lambda}_s\ell)]$ and $[1 - \exp(-\hat{\lambda}_c\ell)]$ are respectively the predicted probability of a single-unit truck and a combination truck being involved in accidents. In general, because accident involvement rates, $\hat{\lambda}_s$ and $\hat{\lambda}_c$, and section length, ℓ, are quite small, the relative risk can be approximated by $[1 - \exp(-\hat{\lambda}_s\ell)]/[1 - \exp(-\hat{\lambda}_c\ell)] \approx \hat{\lambda}_s\ell/\hat{\lambda}_c\ell = \hat{\lambda}_s/\hat{\lambda}_c$. (Note that the above approximation uses the Taylor series expansion that $\exp(-\hat{\lambda}_s\ell) \approx 1 - \hat{\lambda}_s\ell$ and $\exp(-\hat{\lambda}_c\ell) \approx 1 - \hat{\lambda}_c\ell$.)

Accident Involvement Reduction Factor

Using the final model selected for rural Interstate and rural two-lane undivided arterial (i.e., Model 3 in tables 16 and 18), the reduction in the expected number of truck accident involvements and its estimated one standard deviation (from eqs (19) and (20)) due to various improvements in horizontal curvature, vertical grade, and shoulder width of a road section are illustrated in tables 20 through 22. Although these tables show the expected truck accident involvement reductions due to an improvement in one and two geometric design elements, eqs (19) and (20) can be used to estimate the expected reductions due to improvements in any combination of geometric design elements.

Table 20. Expected reductions in truck accident involvements on a rural Interstate road section after an improvement in one geometric design element: examples based on Model 3 in table 16.

Length of Original Curve (mi)	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 12°				
	Reduce 1°	Reduce 2°	Reduce 3°	Reduce 4°	Reduce 5°
0.10	10.6% (±2.5%)	20.1% (±4.5%)	28.6% (±6.0%)	36.2% (±7.2%)	43.0% (±8.1%)
0.25	13.7% (±1.9%)	25.5% (±3.3%)	35.7% (±4.2%)	44.5% (±4.9%)	52.1% (±5.3%)
0.50	18.6% (±2.7%)	33.8% (±4.4%)	46.1% (±5.4%)	56.1% (±5.8%)	64.3% (±6.0%)
0.75	23.2% (±4.3%)	41.1% (±6.6%)	54.8% (±7.7%)	65.3% (±8.0%)	73.4% (±7.8%)
≥1.00	27.6% (±5.8%)	47.6% (±8.6%)	62.1% (±9.6%)	72.5% (±9.5%)	80.1% (±9.0%)

Length of Original Grade (mi)	Vertical Grade (VG): for 2% < VG <9%				
	Reduce 1%	Reduce 2%	Reduce 3%	Reduce 4%	Reduce 5%
0.10	7.8% (±3.1%)	15.0% (±5.7%)	21.6% (±7.9%)	27.7% (±9.7%)	33.4% (±11.3%)
0.50	9.0% (±2.5%)	17.3% (±4.6%)	24.7% (±6.3%)	31.5% (±7.7%)	37.7% (±8.8%)
1.00	10.6% (±2.1%)	20.0% (±3.7%)	28.5% (±5.0%)	36.0% (±5.9%)	42.8% (±6.7%)
≥2.00	13.5% (±2.1%)	25.3% (±3.6%)	35.4% (±4.6%)	44.2% (±5.4%)	51.7% (±5.8%)

	Paved Inside Shoulder Width (ISH) per Direction: for ISH ≤ 12 ft				
	Increase 1 ft	Increase 2 ft	Increase 3 ft	Increase 4 ft	Increase 5 ft
	8.2% (±4.2%)	15.7% (±7.7%)	22.7% (±10.7%)	29.0% (±13.2%)	34.9% (±15.4%)

Notes: Values in parentheses are one standard deviation of the expected reductions above.
1 ft = 0.3048 m; 1 mi = 1.61 km.

Table 21. Expected reductions in truck accident involvements on a rural Interstate road section after an improvement in two geometric design elements: examples based on Model 3 in table 16.

Length of Original Curve (LHC) = 0.10 mi and Length of Original Grade (LVG) = 0.50 mi					
Vertical Grade (VG): for 2% < VG < 9%	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 12°				
	Reduce 1°	Reduce 2°	Reduce 3°	Reduce 4°	Reduce 5°
Reduce 1%	18.7% (±3.1%)	27.3% (±4.4%)	35.0% (±5.6%)	42.0% (±6.6%)	48.1% (±7.4%)
Reduce 2%	26.0% (±4.5%)	33.9% (±5.0%)	40.9% (±5.8%)	47.2% (±6.4%)	52.8% (±7.0%)
Reduce 3%	32.7% (±5.8%)	39.9% (±5.9%)	46.3% (±6.2%)	52.0% (±6.5%)	57.1% (±6.9%)
Reduce 4%	38.8% (±7.0%)	45.3% (±6.7%)	51.1% (±6.7%)	56.3% (±6.7%)	61.0% (±6.9%)
Reduce 5%	44.3% (±7.9%)	50.3% (±7.4%)	55.5% (±7.1%)	60.3% (±7.0%)	64.5% (±6.9%)

Length of Original Curve (LHC) = 0.10 mi					
Paved Inside Shoulder Width per Direction (ISH): for ISH ≤ 12 ft	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 12°				
	Reduce 1°	Reduce 2°	Reduce 3°	Reduce 4°	Reduce 5°
Increase 1 ft	18.0% (±4.4%)	26.7% (±5.3%)	34.5% (±6.3%)	41.4% (±7.1%)	47.6% (±7.8%)
Increase 2 ft	24.7% (±7.2%)	32.7% (±7.3%)	39.9% (±7.5%)	46.2% (±7.9%)	52.0% (±8.2%)
Increase 3 ft	30.9% (±9.8%)	38.2% (±9.3%)	44.8% (±9.0%)	50.7% (±8.9%)	55.9% (±8.9%)
Increase 4 ft	36.6% (±12.0%)	43.3% (±11.1%)	49.3% (±10.5%)	54.7% (±10.0%)	59.5% (±9.7%)
Increase 5 ft	41.8% (±13.8%)	48.0% (±12.7%)	53.5% (±11.8%)	58.4% (±11.1%)	62.8% (±10.5%)

Notes: Values in parentheses are one standard deviation of the expected reductions above.
1 ft = 0.3048 m; 1 mi = 1.61 km.

Table 22. Expected reductions in truck accident involvements on a rural two-lane undivided arterial road section after an improvement in one or two geometric design elements: examples based on Model 3 in table 18.

Length of Original Curve (mi)	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 30°				
	Reduce 1°	Reduce 2°	Reduce 5°	Reduce 10°	Reduce 15°
0.10	9.4% (±1.1%)	18.0% (±2.0%)	39.1% (±3.8%)	62.9% (±4.6%)	77.4% (±4.3%)
0.25	10.0% (±1.8%)	19.0% (±3.3%)	41.0% (±6.1%)	65.2% (±7.4%)	79.5% (±6.8%)
0.50	11.0% (±4.7%)	20.7% (±8.4%)	44.1% (±15.4%)	68.7% (±20.2%)	82.5% (±22.0%)
0.75	11.9% (±7.6%)	22.4% (±13.6%)	47.0% (±26.2%)	71.9% (±42.6%)	85.1% (------)
≥1.00	12.8% (±10.6%)	24.0% (±19.0%)	49.7% (±39.6%)	74.7% (------)	87.3% (------)

	Stabilized Outside Shoulder Width per Direction (OSH): for OSH ≤ 12 ft				
	Increase 1 ft	Increase 2 ft	Increase 3 ft	Increase 4 ft	Increase 5 ft
	3.3% (±1.9%)	6.6% (±3.7%)	9.7% (±5.4%)	12.7% (±6.9%)	15.6% (±8.4%)

Length of Original Curve (LHC) = 0.10 mi					
Stabilized Outside Shoulder Width per Direction (OSH): for OSH ≤ 12 ft	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 30°				
	Reduce 1°	Reduce 2°	Reduce 5°	Reduce 10°	Reduce 15°
Increase 1 ft	12.5% (±1.9%)	20.7% (±2.4%)	41.1% (±3.7%)	64.1% (±4.4%)	78.2% (±4.1%)
Increase 2 ft	15.4% (±3.4%)	23.4% (±3.4%)	43.1% (±3.9%)	65.4% (±4.4%)	78.9% (±4.0%)
Increase 3 ft	18.2% (±4.9%)	26.0% (±4.6%)	45.0% (±4.4%)	66.5% (±4.4%)	79.6% (±3.9%)
Increase 4 ft	21.0% (±6.2%)	28.4% (±5.8%)	46.8% (±5.0%)	67.6% (±4.5%)	80.3% (±3.9%)
Increase 5 ft	23.6% (±7.6%)	30.8% (±6.9%)	48.6% (±5.7%)	68.7% (±4.7%)	81.0% (±3.9%)

Notes: Values in parentheses are one standard deviation of the expected reductions above.
------ Very large standard deviation.
1 ft = 0.3048 m; 1 mi = 1.61 km.

To give a simple illustration of the computation, consider a rural Interstate curved road section i with 0.10 mi (0.16 km) length. By reducing 1 degree per 100-ft (30.48-m) arc of the curve (all else being equal), the expected truck accident involvement reduction percentage and the associated standard deviation are calculated as:

$$
\begin{aligned}
\hat{R}_i &= \{1 - \exp[(x^a_{i,7} - x^b_{i,7})\hat{\beta}_7 + (x^a_{i,13} - x^b_{i,13})\hat{\beta}_{13}]\}\times 100 \\
&= \{1 - \exp[(x^a_{i,7} - x^b_{i,7})\hat{\beta}_7 + (x^a_{i,7}\times x^a_{i,8} - x^b_{i,7}\times x^b_{i,8})\hat{\beta}_{13}]\}\times 100 \\
&= \{1 - \exp[(-1)\times 0.088861 + (-1\times 0.1)\times 0.234209]\}\times 100 \\
&= \{1 - 0.8925\}\times 100 \\
&= 10.6
\end{aligned}
\tag{21}
$$

and

$$
\begin{aligned}
s.d.(\hat{R}_i) \approx\ & \left\{\exp\left[(x^a_{i,7}-x^b_{i,7})\hat{\beta}_7 + (x^a_{i,13}-x^b_{i,13})\hat{\beta}_{13} + \frac{\hat{\tau}}{2}\left((x^a_{i,7}-x^b_{i,7})^2 s_{7,7} + (x^a_{i,13}-x^b_{i,13})^2 s_{13,13} + 2(x^a_{i,7}-x^b_{i,7})(x^a_{i,13}-x^b_{i,13})\hat{\rho}_{7,13}(s_{7,7}s_{13,13})^{1/2}\right)\right]\right\}\times \\
& \left\{\exp\left[\hat{\tau}\left\{(x^a_{i,7}-x^b_{i,7})^2 s_{7,7} + (x^a_{i,13}-x^b_{i,13})^2 s_{13,13} + 2(x^a_{i,7}-x^b_{i,7})(x^a_{i,13}-x^b_{i,13})\hat{\rho}_{7,13}(s_{7,7}s_{13,13})^{1/2}\right)\right] - 1\right\}^{1/2}\times 100 \\
=\ & \left\{\exp\left[-0.1123 + \frac{1.57}{2}((-1)^2(0.028)^2 + (-1\times 0.1)^2(0.084)^2 + 2\times(-1)(-1\times 0.1)(-0.792)(0.028)(0.084))\right]\right\}\times \\
& \{\exp[1.57((-1)^2(0.028)^2 + (-1\times 0.1)^2(0.084)^2 + 2\times(-1)(-1\times 0.1)(-0.792)(0.028)(0.084))] - 1\}^{1/2}\times 100 \\
=\ & 2.5
\end{aligned}
\tag{22}
$$

where $(s_{7,7})^{1/2}$ and $(s_{13,13})^{1/2}$ are the standard deviations of the estimated regression parameters $\hat{\beta}_7$ and $\hat{\beta}_{13}$, respectively, and are available in table 16; $\hat{\tau} = 1.57$; and $\hat{\rho}_{7,13} = -0.792$ (table 19).

The Poisson regression model introduced in this report can be developed and tested for other States in a similar manner. For those States where detailed roadway and accident data are not available for conducting such an analysis, we recommend that the models developed in this study be used with a slight modification. For example, for truck accident involvement rate on rural Interstate the modification can be made as follows:

$$
\begin{aligned}
\hat{\lambda}_i = \left[\frac{AR}{0.81}\right] \exp(& -0.626471 + 0.0244x_{i,6} + 0.088861x_{i,7} + 0.234209x_{i,13} \\
& + 0.077815x_{i,9} + 0.033973x_{i,14} + 0.085763x_{i,11} - 0.025233x_{i,12})
\end{aligned}
\tag{22}
$$

where 0.81 is the overall truck accident involvement rate per MTM for the rural Interstate road sections examined in this study, and AR represents the overall truck accident involvement rate per MTM for the rural Interstate highways in another State of interest. This modification is intended to adjust for the differences between Utah and the State of interest in, e.g., weather and socioeconomic conditions, as well as the differences in accident reporting practices for nonfatal accidents and in the criteria used for classifying roadways. Note that the expected number of truck accident involvements is computed as $\hat{\mu}_i = v_i\hat{\lambda}_i$. Under this modified model, the expected percentage reductions in truck accident involvements and the associated standard deviations can still be computed from eqs (19) and (20) without any changes. Thus, the expected percentage reductions presented in this section can still be used in other States. Again, it is recommended that these estimates from the modified models be used with some caution. In addition, any attempt to extrapolate these rates outside of the range of the variables, as specified in tables 13 through 15, is not recommended.

To give an example of how truck accident involvement rate and accident probability can be computed from eq (22) for rural Interstate highways in another State, let us consider one hypothetical road section with the following characteristics:

(1) Lane width: 12 ft; (2) Section length: 0.3 mile; (3) Number of lanes: 4;
(4) AADT/lane: 2,500 vehicles/lane; (5) Horizontal curvature: 3 degrees/100-ft arc;
(6) Length of original curve: 0.5 mi; (7) Vertical grade: 3 percent;
(8) Length of original grade: 0.3 mi; (9) Pave outside shoulder width: 10 ft;
(10) Paved inside shoulder width: 6 ft; and (11) Percent trucks: 20.

Also, assume that the overall truck accident involvement rate on rural Interstate in that State is 1.25 trucks per MTM (i.e., $AR=1.25$). First, truck exposure in a year can be computed as $v_i=365\times(\text{AADT})\times(\text{percent trucks}/100)\times(\text{section length})=365\times(4\times2{,}500)\times(20/100)\times0.3=0.219$ MTM. Second, based on eq (22), truck accident involvement rate is estimated as:

$$\begin{aligned}\hat{\lambda}_i &= [1.25/0.81]\times\exp\{-0.626471+0.0244\times(2500/1000)+0.088861\times3+0.234209\times(3\times0.5)\\&\quad+0.077815\times3+0.033973\times(3\times0.3)+0.085763\times(12-6)-0.025233\times20\}\\&= 1.543\times\exp\{0.32636\} = 2.14 \text{ (trucks per MTM).}\end{aligned}$$

The expected number of truck accident involvements in a 1-year period is estimated at $2.14\times0.219\approx0.47$ trucks. The probability of observing "y" trucks involved in accidents on this particular road section in 1 year is then $p(y)=[(0.47)^y\exp(-0.47)]/y!$. For example, the probability of observing one truck involved in accidents is $p(y=1)=[(0.47)^1\exp(-0.47)]/1!=0.29$.

SUMMARY AND FUTURE RESEARCH

The Poisson regression model was employed to establish empirical relationships between truck accidents and key highway geometric design elements. For a particular roadway type, the number of trucks involved in accidents on each road section over a 1-year period was assumed to be Poisson distributed, and the Poisson rate (in number of trucks involved in accidents per truck miles traveled) was related to highway geometric, traffic, and other potential explanatory variables by a loglinear function. Both the uncertainties in truck exposure data and the omitted variables in the model could create extra variations (or overdispersion) in the Poisson regression model. Covariance and t-statistic of the estimated parameters from the Poisson regression model were then adjusted to reflect this possible overdispersion.

The primary source of data used in this study was the HSIS. Highway geometric and traffic data for three roadway types: rural Interstate, urban Interstate and freeway, and rural two-lane undivided arterial, and the associated truck accidents for Utah from 1985 to 1989 were used for analysis. Limited by the available number of truck accidents, one model was developed for all large trucks and accident severity types combined. Furthermore, because the truck exposure data are currently not available by time of day and weather conditions, the developed models are not capable of distinguishing the effect of highway geometric design on the safety performance of trucks between daytime and nighttime and between different weather conditions.

Despite the limitations in existing Utah geometric design data, e.g., lack of variation in lane width, some encouraging relationships were developed for horizontal curvature, vertical grade, and shoulder width. To some extent, the final models selected for the three roadway types all failed to pass the chi-square test at a 5-percent α level, especially the one for urban Interstate and freeway. This indicates that the overdispersion problem, due most likely to the uncertainties in truck exposure data and the omitted variables in the models, was quite significant, and that the model can probably be improved by including additional explanatory variables, by separating daytime and nighttime travel, and by improving the accuracy of truck exposure data. The effect of correcting for the overdispersion was found to lower the significance level of the estimated regression parameters. However, it did not significantly alter the conclusions reached regarding the relationships between truck accidents and the examined traffic and highway geometric variables. In the next chapter, an alternative model is used to quantify the contribution of the uncertainty on truck exposure data and the omitted variables to the overdispersion in the model.

Some recommended extensions of this study are as follows:

1. Apply the proposed model framework to other roadway types and consider truck accidents by truck type. The latter will require the truck exposure data to be available by truck type for individual road sections.

2. Develop a unified statistical model framework to establish the relationships between highway geometric design and truck accident frequency, accident type, and accident severity simultaneously. This perhaps could be achieved through the use of some joint probability distributions.

3. Broaden the data base to include other weather, terrain, and socioeconomic conditions.

4. Study the effect of continuous geometric design conditions, e.g., sharp horizontal curves following long segments of generally straight alignment and contiguous horizontal curves. This requires the information on the direction of travel at the time of accidents.

5. Develop an integrated model that is capable of describing the effects of travel lane and shoulder design, as well as the roadside design, on truck accidents.

5. A NEGATIVE BINOMIAL REGRESSION MODEL

INTRODUCTION

The model described in the previous chapter postulates that the number of trucks involved in accidents on each road section over a period of time follows a Poisson distribution, and the rate of the Poisson is related to highway geometric and traffic variables by a loglinear function. The developed Poisson regression models were tested using Pearson's chi-square test. It was found that all three final models (one for each roadway type) failed the test at a 5-percent α level, indicating that the variance of the accident data was greater than what the model predicted. In other words, extra variations (or overdispersion) over the Poisson models exist in the data. To be specific, there are some variations in the truck accident data that were not explainable by the Poisson model using the available explanatory variables.

The overdispersion could come from several possible sources. Two sources were identified as the primary contributors to the overdispersion in the Poisson models: the uncertainty on truck exposure (or travel) data and the "omitted variables." The former is due to the fact that because truck exposure data were estimated using the data collected by a sampling system (HPMS), the data were subject to sampling errors as well as nonsampling errors. The latter recognizes the fact that there were only a limited number of variables available for explaining the variations in accident data. A third possible source is that the highway environment is not homogeneous within each sampling period, e.g., truck accident involvement rate during daytime and nighttime might be different. A fourth possible source is that the occurrences of accidents on different road sections might be positively correlated, rather than independent.[71]

To investigate the effects of the uncertainty in truck exposure data and the omitted variables on the overdispersion of the developed Poisson models, it is first shown statistically how these two factors affect the Poisson model. Then, a negative binomial based regression model, which allows overdispersion in the model, is used to quantify the effects. The data used for developing the Poisson models are reinvestigated using the negative binomial regression. The results are compared to those obtained from the Poisson regression models.

This chapter is organized as follows. First, a review of the uncertainties associated with truck exposure data under HPMS is presented. Second, it is shown how the uncertainty on truck exposure data and the omitted variables affect the Poisson model, and how the negative binomial regression model could be used to quantify these effects. Third, the results from the negative binomial regression models are presented and compared with those obtained using the Poisson regression models. The final section concludes the chapter.

UNCERTAINTIES ON TRUCK EXPOSURE DATA FROM THE HPMS

The statistical validity and implementation procedures of HPMS are well documented in FHWA's Highway Performance Monitoring System Field Manual and the Traffic Monitoring Guide (TMG).[46,56] In addition, a comprehensive statistical analysis addressing potential sources of sampling errors associated with traffic volume and vehicle classification counts was given in Hallenbeck and Bowman (H-B).[72] In sum, road sections are stratified by area (rural vs. urban), functional class (not by roadway type), and traffic volume, and the statistical sampling method used is stratified random sampling.

For a sample road section *i* in a year, the annual vehicle miles of travel (VMT) of a particular truck type of interest is estimated as:

$$V_i = 365 \times AADT_i \times \ell_i \times T\%_i / 100 \tag{23}$$

where $T\%_i$ is the estimated percent trucks and ℓ_i is the section length in miles. By assuming that $AADT_i$ and $T\%_i$ are uncorrelated, the variance of the truck travel V_i is:

$$Var(V_i) = (\phi_a^2 + \phi_b^2)[E(V_i)]^2 = \phi^2 [E(V_i)]^2 \tag{24}$$

where ϕ_a, ϕ_b, and ϕ are respectively the coefficients of variations (CV's) of AADT, percent trucks, and truck VMT estimates of the section.

For a non-HPMS road section, AADT and percent trucks can be estimated from adjacent HPMS sampled sections, which are preferably located along the same route. However, the uncertainties in the estimates for non-HPMS sample road sections are expected to be slightly higher than those of the sampled sections. The CV's of AADT, percent trucks, and truck VMT estimates (i.e., ϕ_a, ϕ_b, and ϕ) are estimated in the remainder of the section.

Uncertainties on Vehicle Volume (AADT)

A 48-hour monitoring period for vehicle volume, vehicle classification, and truck weight monitoring was recommended in the TMG. For an HPMS sample road section *i* during a year, the AADT estimate, $AADT_i$, is obtained by adjusting the collected 48-hour axle counts with seasonal, day-of-week, and axle correction factors if the section is sampled, and with additional adjustment for growth factor if the section is not currently sampled. The equation used to estimate AADT is:

$$AADT_i = (0.5 \times NAXL_i) \times SEA \times WK \times \frac{1}{ACF} \times GROW \qquad (25)$$

where NAXL is the total number of axle counts collected from the road section during a 48-hour monitoring period; SEA and WK are respectively seasonal (or monthly) and day-of-week adjustment factors; ACF is an axle correction factor, defined as average number of axles per vehicle per day; and GROW is an annual growth factor. All the adjustment factors are estimated from historical data by highway functional class and area type, and the proposed procedures for developing these factors are described in TMG. Assuming that the standard deviation of each factor in eq (25) is a small fraction of its mean and that all the factors are independent, an approximation of the variance of $AADT_i$ can be shown to be:[56,72]

$$\begin{aligned} Var(AADT_i) &= \left[CV_v^2 + CV_s^2 + CV_w^2 + CV_a^2 + CV_g^2\right]\left[E(AADT_i)\right]^2 \\ &= \phi_a^2 \left[E(AADT_i)\right]^2 \end{aligned} \qquad (26)$$

where CV_v, CV_s, CV_w, CV_a, and CV_g are, respectively, coefficient of variations of the 48-hour axle volume counts across days, seasonal adjustment factor, day-of-week adjustment factor, axle correction factor, and annual growth factor. Some example CV's were obtained by H-B based on the HPMS data from five participating States (table 23).[72] CV_v and CV_a are the two major sources of variations. For example, the standard deviation of estimated AADT for a rural Interstate sample section is roughly 20 percent (i.e., $\phi_a = (0.117^2 + 0.05^2 + 0.140^2 + 0.02^2)^{1/2}$) of the expected value of AADT.

Uncertainties on Vehicle Classification

Ideally, the uncertainties of estimated percent trucks $T\%_i$, can be estimated in a similar manner as that of AADT. Limited resources, however, prohibit a State from collecting vehicle classification data for all HPMS sample road sections and from determining the CV's of adjustment factors for different vehicle types and highway functional classes directly.[72] For this reason, TMG recommends that road sections selected for collecting vehicle classification data be a subset of the road sections selected for collecting vehicle volume data using a simple random selection procedure within each stratum. Furthermore, TMG recommends that 300 measurements

Table 23. Coefficient of variations (CV's) of vehicle volume, percent trucks, and truck VMT.

	AADT Adjustment Factors				AADT	Percent Trucks A Highway Class		Percent Trucks A Road Section		Truck VMT		
						CV_h		ϕ_b		ϕ		
Roadway Type	CV_v	CV_a	CV_s	CV_g	ϕ_a	SU	CB	SU	CB	SU	CB	ALL
Rural Interstate	0.117	0.140	0.05	0.03	0.20	0.12	0.15	0.18	0.22	0.27	0.30	0.29
Urban Interstate & Freeway	0.078	0.067	0.03	0.03	0.11	0.19	0.18	0.29	0.27	0.31	0.29	0.30
Urban Multilane Divided Arterial	0.069	0.058	0.03	0.03	0.10	0.18	0.19	0.26	0.29	0.28	0.31	0.29
Rural Two-Lane Undivided Arterial	0.078	0.067	0.05	0.02	0.12	0.25	0.19	0.38	0.29	0.40	0.31	0.34

Notes:
(1) CV_v: CV of daily vehicle count; CV_a: CV of axle correction factor; CV_s: CV of seasonal factor; CV_g: CV of annual growth factor; ϕ_a: CV of AADT; CV_h: CV of percent trucks for an entire highway class; ϕ_b: CV of percent trucks of an individual road section; ϕ: CV of truck VMT; SU: single-unit trucks; CB: combination trucks; ALL: all large trucks.
(2) CV_v and CV_a are obtained from reference 72 (Exhibit A-18, pp A.22).
(3) CV_s and CV_g are assumed values.
(4) ϕ_a is estimated using eq (26), ignoring the CV of day-of-week factor.
(5) ϕ_b is estimated from Hallenbeck and Bowman and Kansas State DOT with some additional assumptions.[72,73]
(6) ϕ's for SU and CB are estimated using eq (24); and ϕ for all trucks is computed as a weighted average of that of SU and CB, in which truck exposures of the two truck types are used as the weights.

be taken over a 3-year cycle (i.e., 100 measurements each year) for 9 highway groups. The 9 highway groups are stratified by functional class and area type as: (1) rural Interstate, (2) rural other principal arterial, (3) rural minor arterial, (4) rural collector, (5) urban Interstate, (6) urban other freeway and expressway, (7) urban other principal arterial, (8) urban minor arterial, and (9) urban collector. It is also recommended that these measurements be spread out systematically throughout the year to capture seasonal variations. Following this TMG guideline, after a 3-year cycle, a State would have approximately 30 vehicle classification measurements for each highway group to estimate the vehicle percentages by vehicle type. Since the road sections selected to collect vehicle classification data are a subset of the road sections selected to collect vehicle volume data, the vehicle distribution by type for vehicle volume sample sections that are not vehicle classification sample sections are estimated by assigning the stratum averages from the specific stratum.

For each highway group, the overall precision level for the estimated vehicle percentage by vehicle type is given in TMG and H-B as:

$$d_k^2 = \frac{z_c^2 (\mathrm{TCV}_k)^2}{nc} \tag{27}$$

where d_k is the accuracy (or relative error) of type k vehicle percentage as a fraction of its estimate, nc is the number of counts (or measurements) in the highway group, TCV_k is the total coefficient of variation for the percentage of type k vehicles in the entire highway group during the year, including daily, locational, and seasonal variations, and z_c is a standard normal deviate of probability C, e.g., $z_{0.95} = 1.96$. Equivalently, the standard deviation of type k vehicle percentage estimate is d_k/z_c ($=\mathrm{TCV}_k/(nc)^{1/2}$). Based on the five-state data in H-B, a typical State with 30 vehicle classification counts in each highway group, the sampling CV values for the estimated vehicle percentage for 2 truck types and 4 roadway types are given in table 23. Note that the TCV's of three-axle single-unit and 3S2 trucks (truck trailers or tractor semitrailers with five axles) in H-B (listed in the exhibit A-1 of H-B) are used respectively to represent the TCV's of single-unit and combination trucks.

The CV's of the estimated percent trucks are for a specific highway group. For an individual HPMS section, its corresponding CV's are expected to be higher than those for a specific highway group. That is, when we consider the uncertainties of the estimated percent trucks for an individual road section, we would expect it to be higher than those for the system. In addition to the sampling error described above, estimated percent trucks is also subjected to measurement error (or misclassification error), resulting from the use of automatic recording machines. Field tests conducted by the Kansas State Department of Transportation for the FHWA revealed that vehicle classifiers could over- or under-count vehicles of any type, and the tested vehicle classifier over-counted about 17 percent and 8.8 percent of three-axle single-unit and combination trucks, respectively.[73] The Kansas study also found that the tested system had problems with slow-moving vehicles (less than 20 mi/h or 32.2 km/h) and vehicles in queues. No statistical studies are available to the researchers for further assessments of the effects of the measurement error. However, the overall performance of the vehicle classification equipment was found to have improved significantly over the years. For simplicity, we assume that the estimated percent trucks of each HPMS sample section is unbiased and its CV is 1.5 times the CV for the highway group to which the road section belongs. Based on this assumption, table 23 also gives CV's of the estimated percent trucks of a specific road section. Using eq (24), the CV's of estimated truck exposure are estimated and presented in table 23. For example, for a rural Interstate road section, the CV of estimated combination truck VMT is approximately 0.30 ($\phi=(0.20^2+0.22^2)^{1/2}$).

THE PROPOSED NEGATIVE BINOMIAL REGRESSION MODEL

A negative binomial regression model, which allowed overdispersion in the model, was developed to quantify the effect of uncertainty on truck exposure data and the effect of the omitted variables on the overdispersion of the Poisson models developed in chapter 4.

Model Formulation

Consider a set of n highway sections of a particular roadway type, say, rural Interstate. Let Y_i be a random variable representing the number of trucks involved in accidents on highway section i during a 1-year period. Further, assume that the amount of truck travel or truck exposure on this highway section, V_i, is also a random variable, estimated through a highway sampling system, such as the HPMS. Associated with each highway section i, there is a $k \times 1$ vector of explanatory variables, denoted by $\underline{x}_i = (x_{i1}=1, x_{i2}, \ldots, x_{ik})'$, describing its geometric characteristics, traffic conditions, and other relevant attributes. Given V_i and $\underline{x}_i$, truck accident involvements Y_i, $i=1,2,\ldots,n$, are postulated to be independent, and each is Poisson distributed as:

$$p(Y_i = y_i \mid \Lambda_i = \tilde{\lambda}_i, V_i = \tilde{v}_i, x_i) = \frac{(\tilde{\lambda}_i \tilde{v}_i)^{y_i} e^{-\tilde{\lambda}_i \tilde{v}_i}}{y_i!}, \qquad (i = 1,2,\ldots,n;\ y_i = 0,1,2,\ldots) \tag{28}$$

where Λ_i (>0) is the truck accident involvement rate on highway section i, and its distribution is expected to vary from one highway section to another, depending on its explanatory variables $\underline{x}_i$. For each highway section i, the Poisson model implies that the *conditional mean* equals *conditional variance.*

To establish a relationship between truck accident involvement rate and highway geometric and traffic variables, the loglinear rate function $\Lambda_i = \exp(\underline{x}_i'\underline{\beta} + \varepsilon_i)$ is used in this study, where $\underline{\beta}$ is a $k \times 1$ parameter vector, and ε_i is a specification error due, for instance, to "omitted variables" which are independent of $\underline{x}_i$. In cases where $\underline{x}_i$ and V_i are given with no (or negligible) uncertainties and Λ_i is assumed to be a constant (i.e., ε_i=0, for all i), eq (28) becomes a classical Poisson regression model. Without loss of generality, one can assume that $E(\exp(\varepsilon_i))=1$ and $Var(\exp(\varepsilon_i))=\eta^2$ (see, e.g., Dean and Lawless).[60]

It is reasonable to assume that truck exposure V_i and model specification error ε_i are independent. Given the explanatory variables $\underline{x}_i$, the mean and variance of $M_i=\Lambda_i V_i$ can be easily shown to be:

$$\begin{aligned} E(\Lambda_i V_i \mid x_i) &= \lambda_i v_i = \mu_i \\ Var(\Lambda_i V_i \mid x_i) &= [(\eta^2+1)(\phi^2+1) - 1](\lambda_i v_i)^2 = \theta \mu_i^2 \end{aligned} \tag{29}$$

where λ_i and v_i are respectively the expected value of Λ_i and V_i; and ϕ is the coefficient of variation of the estimated truck exposure which is approximately 0.3 under the HPMS (see table 23). Therefore, the mean and variance of Y_i, conditional on $\underline{x}_i$, are:

$$\begin{aligned} E(Y_i \mid x_i) &= E\left[E(Y_i \mid \varepsilon_i, V_i, x_i)\right] = \mu_i \\ Var(Y_i \mid x_i) &= E\left[Var(Y_i \mid \varepsilon_i, V_i, x_i)\right] + Var\left[E(Y_i \mid \varepsilon_i, V_i, x_i)\right] \\ &= \mu_i + \theta \mu_i^2 \end{aligned} \tag{30}$$

where θ is called a dispersion parameter. Extra variance over the Poisson model, $\theta\mu_i^2$, is introduced into the Poisson model as a result of the additional uncertainties on the estimated truck exposure and the "omitted variable" error ε_i.

The same mean and variance function as those in eq (30) can be derived if we made a stronger assumption that random variable M_i (= $\Lambda_i V_i$) follows a gamma distribution. That is, M_i~Gamma(μ_i,α_i) with density, mean, and variance:

$$p_{M_i}(m_i) = \frac{1}{\Gamma(\alpha_i)}\left(\frac{\alpha_i m_i}{\mu_i}\right)^{\alpha_i} \exp\left(\frac{-\alpha_i m_i}{\mu_i}\right)\frac{1}{m_i} \tag{31}$$

$$E(M_i) = \mu_i \quad \text{and} \quad Var(M_i) = \frac{1}{\alpha_i}\mu_i^2$$

where $m_i > 0$, $\mu_i > 0$, and $\alpha_i > 0$, and μ_i is the mean and α_i is the precision parameter. This assumption would make Y_i, given $\underline{x}_i$, negative binomial distributed as (see, e.g., Cameron and Trivedi):[64]

$$p(Y_i = y_i | \underline{x}_i) = \frac{\Gamma(y_i + \alpha_i)}{\Gamma(y_i + 1)\Gamma(\alpha_i)}\left(\frac{\alpha_i}{\alpha_i + \mu_i}\right)^{\alpha_i}\left(\frac{\mu_i}{\alpha_i + \mu_i}\right)^{y_i} \tag{32}$$

By assuming a constant precision parameter $\alpha_i = 1/\theta$, the negative binomial distribution in eq (32), conditional on explanatory variables, has mean and variance exactly the same as those in eq (30). Note that the Poisson distribution can be regarded as a limiting distribution of the negative binomial distribution as α_i approaches ∞ or θ approaches 0.[74]

Although the discussion above does not distinguish accidents by truck configuration and accident severity, in principal, the overall framework could be applied to any truck type and accident severity type of interest, provided that there are enough accident data and that truck exposure by truck type can be properly estimated.

Test Statistics and Estimation Procedures

As indicated in the last chapter, the consequences of ignoring the overdispersion in the Poisson model are that consistent estimates of $\underline{\beta}$ under the Poisson model (such as maximum likelihood estimates (MLE) or the quasi-likelihood function estimates of McCullagh and Nelder) are still consistent; however, the variances of the estimated parameters are not.[58] Consequently, the test statistics, such as asymptotic t-statistics, are not consistent. Cameron and Trivedi's study suggested that the estimates from the Poisson model and those based on negative binomial distributions are quite similar, especially, for those parameters with high t-ratios, but the variances of the estimated parameters under Poisson assumption are generally substantially underestimated.[64] The underestimation is the result of restraining the conditional variance of the data to be equal to the conditional mean.

Several test statistics for testing H_0: $\theta = 0$ against H_A: $\theta > 0$, have been proposed for the particular mean-variance relationship in eq (30). Test statistics based on the null Poisson model were investigated by Cox, Cameron and Trivedi, and Lee.[59, 64, 65, 75] The following score test statistic was proposed:

$$T_{score} = \frac{\sum_{i=1}^{n}\left[(y_i - \hat{\mu}_i)^2 - y_i\right]}{\left[2\sum_{i=1}^{n}\hat{\mu}_i^2\right]^{1/2}} \tag{33}$$

where $\hat{\mu}_i = v_i \exp(\underline{x}_i' \hat{\underline{\beta}})$, in which $\hat{\underline{\beta}}$ is the MLE for $\underline{\beta}$ in the Poisson model. Under the null hypothesis H_0, T_{score} is asymptotically distributed as a standard normal as n increases. That is, $T_{score} \rightarrow N(0,1)$ as n increases. Dean and Lawless have refined this score test statistic for small sample conditions.[60] These tests are local score tests computed only under the null hypothesis and avoid the more burdensome computation of estimates under the alternative hypothesis. However, once the Poisson model (or the null hypothesis) is rejected, these tests do not automatically lend themselves for estimating both θ and $\underline{\beta}$. Another estimation procedure has to be used.

A straightforward way of handling this overdispersion problem is to accept the negative binomial distribution of eq (32) and to proceed with classical estimation and testing procedure, based largely on asymptotic theory (e.g., Collings and Margolin, Cameron and Trivedi).[76,64] The MLE and the associated t-tests of parameters using negative binomial distribution have been given in Lawless.[74] Although, the negative binomial model is more general than the Poisson model, it requires much more extensive numerical computation to estimate model parameters and to generate inferential statistics. In addition, the statistical properties of different estimators, e.g., MLE and moment estimators, of the negative binomial regression model for this particular problem have not yet been fully investigated.

RESULTS

The final models developed in the last chapter are reinvestigated in this section using the negative binomial assumption. The maximum likelihood estimates of the regression parameters (including the dispersion parameter θ), score test statistics, and loglikelihood function evaluated at the MLE are presented in table 24 for the three roadway classes examined in this study. The estimated parameters of the Poisson regression models obtained in the last chapter, together with the adjusted asymptotic standard deviations and t-statistics, are also given in the table for comparison purpose. Some observations are made from the table as follows:

1. For rural Interstate and rural two-lane undivided arterial, the regression parameters derived from the negative binomial regression are similar to those obtained from the Poisson regression. In most cases, the adjusted t-statistics of the Poisson models are consistent with the t-statistics obtained from the negative binomial models. On the other hand, for urban Interstate and freeway some differences in the estimated regression parameters can be observed between the two models for horizontal curvature, the interactive term of horizontal curvature and length of original curve, paved inside shoulder width, and number of lanes. For example, the estimated parameter for the deviation of paved inside shoulder width from the ideal shoulder width of 12 ft (3.66 m) is 0.1539 ($\hat{\beta}_{11}$) in the Poisson model, while it is 0.0832 under the negative binomial model. Based on the loglikelihood function and *AIC* values, the negative binomial models for all three roadway classes are consistently better than their corresponding Poisson models, especially the one for urban Interstate and freeway.

2. The overdispersions are estimated to be statistically significant for all roadway types, according to both the score test of the Poisson models and the estimated dispersion parameter, $\hat{\theta}$, from the negative binomial regression models. Specifically, all the T_{score} statistics are substantially greater than 1.96, and the t-statistics of the estimated dispersion parameters are highly significant. The estimated dispersion parameters, $\hat{\theta}$, and the associated t-statistics (in parentheses) for the three roadway types are respectively: 0.95 (8.89), 0.58 (9.07), and 1.22 (5.73).

4. Using the estimated dispersion parameters (i.e., θ=0.95, 0.58, and 1.22) and the coefficient of variations on truck VMT (i.e., ϕ =0.29, 0.30, and 0.34 in table 23), the extra variations contributed by the omitted variables for the three roadway types using eq (29) can be estimated as follows: η^2 = 0.80, 0.45, and 0.99. (Note that ϕ^2 = 0.084, 0.090, and 0.116 for the three roadway classes.) Comparatively, the omitted variables contributed to the overdispersion of the Poisson model 5 to 10 times more than the uncertainty on truck exposure did. These extra variations can most likely be reduced by including more explanatory variables, such as unpaved shoulder width and type, superelevation, and roadside design, or more detailed truck exposure information by truck type, by daytime and nighttime, and by weather conditions.

5. One interesting observation from table 24 is that the MLE of the negative binomial regression models tend to overpredict the total truck accident involvements, while the Poisson estimates are essentially right on target! The negative binomial models overpredict the total truck accident involvements by 3.6, 7.5, and 6.9 percent for rural Interstate, urban Interstate and freeway, and rural two-lane undivided arterial, respectively. However, these amounts of overprediction do not seem to be unreasonable, given the stochastic nature of the vehicle accident process. In situations where the MLE of the negative binomial model substantially overpredicts or underpredicts total accident involvements, one may want to consider a constrained MLE which sets the predicted total truck accident involvements to be close to the observed total. Other statistical estimators, such as the moment estimator, could also be considered.

6. The final selected model for urban Interstate and freeway is the negative binomial regression model in table 24, which include the following variables: AADT per lane, number of lanes, horizontal curvature, length of original curve, vertical grade, paved inside shoulder width, and percent trucks. The truck accident involvement rate for the final selected model is computed as: $\hat{\lambda}_i = \exp[\hat{\beta}_1 + (\hat{\beta}_3+\hat{\beta}_4+\hat{\beta}_5)/3 + x_{i,16}\hat{\beta}_{16} + x_{i,6}\hat{\beta}_6 + x_{i,17}\hat{\beta}_{17} + x_{i,7}\hat{\beta}_7 + x_{i,13}\hat{\beta}_{13} + x_{i,9}\hat{\beta}_9 + x_{i,11}\hat{\beta}_{11} + x_{i,12}\hat{\beta}_{12}] = \exp[-0.559480 + 1.30802x_{i,16} + 0.050161x_{i,6} + 0.088493x_{i,17} + 0.053897x_{i,7} + 0.049554x_{i,13} + 0.093379x_{i,9} + 0.083181x_{i,11} - 0.084985x_{i,12}]$.

For the ranges of variables indicated in table 14, the final model suggests the following relationships between traffic/geometric design variables and truck accident involvement rates (based on eqs (19) and (20) in chapter 4):

(1) As AADT per lane increases by 1,000 vehicles per lane, truck accident involvement rate increases by 5.1%.

(2) As horizontal curvature increases, truck accident involvement rate increases. However, the increase depends on the length of curve. For example, for a curve with 0.1 mi (0.16 km) in length and with curvature greater than 1 degree per 100-ft (30.48-m) arc, as horizontal curvature increases by 1 degree, truck accident involvement rate increases by about 6.0 percent.

(3) As vertical grade increases by 1 percent, truck accident involvement rate increases by 9.8 percent.

(4) As the length of curve increases, truck accident involvement rate increases. The increase, however, depends on the curvature degree. For example, for a 3-degree

curve, as the length of curve increases by 0.1 mi (0.16 km), truck accident involvement rate increases by about 1.5 percent.

(5) As the paved inside shoulder width increases by 1 ft (0.3048 m) per direction, truck accident involvement rate decreases by about 8.0 percent.

(6) *For a constant vehicle density*, as percent trucks in the traffic stream increases by 1 percent, truck accident involvement rate decreases by about 8.1 percent.

(7) To illustrate the effect of "number of lanes," consider two road sections with the same traffic density (i.e., the same value in AADT per lane) and geometric design conditions, and one has nl lanes and the other has $nl+2$ lanes (where $4 \leq nl \leq 6$). The model suggests that the section with $nl+2$ lanes is expected to have a 19.4 percent higher truck accident involvement rate than the one with nl lanes. As indicated in the last chapter, this is likely due to the increased frequency in lane changing and overtaking movements as number of lanes increases.

To give another example on the effect of number of lanes, consider one 4-lane road section with AADT per lane equal to 15,000 vehicles. Now, if the road section is expanded to 6 lanes (all else being the same), we have AADT per lane reduced from 15,000 to 10,000 vehicles (and the number of lanes increased by 2). For this case, the model suggests that truck accident involvement rate would decrease by about 7.1 percent. Thus, in this particular case the decrease in truck accident involvement rate due to reduced traffic density is able to offset the increase in truck accident involvement rate due to the increase in number of lanes.

(8) In Utah, for the same geometric design and traffic conditions, urban freeways had a considerably higher overall truck accident involvement rate than urban Interstate highways ($\hat{\beta}_{16} = 1.3080$).

7. Using the negative binomial regression model developed for urban Interstate and freeway, the reduction in the expected number of truck accident involvements and its estimated one standard deviation due to various improvements in horizontal curvature, vertical grade, and paved inside shoulder width of a road section are illustrated in table 25 and table 26.

SUMMARY

The negative binomial regression models developed in this chapter indicate that there is a significant amount of overdispersion in the accident data over the estimated Poisson regression model developed in chapter 4. These extra variations can most likely be reduced by including more explanatory variables, e.g., by including additional variables such as roadside design and superelevation, by including more detailed truck exposure data by truck type, by time of day, and by weather conditions, and by improving the quality of the truck exposure data. However, the results reaffirm that the conclusions reached by using the Poisson regression regarding the relationships between truck accidents and the examined traffic and highway geometric design variables for rural Interstate and rural two-lane undivided arterial were valid even when the overdispersion exists in the models. Although, the negative binomial model is more general than the Poisson model, it requires more extensive numerical computation to estimate model parameters and to generate inferential statistics. In addition, the statistical properties of different estimators, e.g., MLE and moment estimators, of the negative binomial regression model for this particular problem have not yet been fully investigated.

Table 24. Estimated parameters of the tested Poisson and negative binomial regression models and associated statistics.

Roadway Class (No. of Sections)	Rural Interstate (8,263 Sections)		Urban Interstate & Freeway (2,810 Sections)		Rural 2-Ln Undivided Arterial (13,634 Sections)	
Model	Poisson	Negative Binomial	Poisson	Negative Binomial	Poisson	Negative Binomial
β_1 Dummy intercept	-0.431762 (±0.360;-1.20)	-0.265214 (±0.349;-0.76)	-0.947077 (±0.665;-1.42)	-0.221901 (±0.496;-0.45)	0.434863 (±0.277;1.57)	0.463495 (±0.228;2.04)
β_2 Dummy variable for 1986	-0.183853 (±0.108;-1.71)	-0.204387 (±0.104;-1.96)	-0.385215 (±0.125;-3.08)	-0.389702 (±0.102;-3.85)	-0.181743 (±0.148;-1.23)	-0.221259 (±0.126;-1.75)
β_3 Dummy variable for 1987	-0.161461 (±0.106;-1.52)	-0.139613 (±0.104;-1.35)	-0.582372 (±0.130;-4.51)	-0.551033 (±0.103;-5.34)	-0.461769 (±0.159;-2.90)	-0.528860 (±0.134;-3.93)
β_4 Dummy variable for 1988	-0.111511 (±0.106;-1.05)	-0.083996 (±0.104;-0.80)	-0.292152 (±0.118;-2.48)	-0.231267 (±0.096;-2.39)	-0.344109 (±0.147;-2.33)	-0.398717 (±0.124;-3.21)
β_5 Dummy variable for 1989	-0.311155 (±0.110;-2.83)	-0.311454 (±0.108;-2.90)	-0.273012 (±0.115;-2.39)	-0.230436 (±0.095;-2.43)	-0.253417 (±0.144;-1.76)	-0.298484 (±0.122;-2.45)
β_{16} Dummy variable for urban freeways	-----	-----	1.40603 (±0.210;6.68)	1.30802 (±0.161;8.14)	-----	-----
β_6 AADT per lane (10^3)	0.024400 (±0.019;1.27)	0.024621 (±0.020;1.22)	0.046010 (±0.010;4.75)	0.050161 (±0.008;6.30)	0.102226 (±0.050;2.03)	0.143453 (±0.045;3.17)
β_{17} Number of lanes (4 to 8 lanes)	-----	-----	0.124950 (±0.053;2.34)	0.088493 (±0.040;2.23)	-----	-----
β_7 Horizontal curvature	0.088861 (±0.035;2.51)	0.073650 (±0.032;2.31)	0.016375 (±0.062;0.26)	0.053897 (±0.044;1.24)	0.094931 (±0.020;4.53)	0.094112 (±0.016;5.80)
β_{13} Horizontal curvature × Length of original curve	0.234209 (±0.105;2.22)	0.277068 (±0.100;2.77)	0.128738 (±0.152;0.85)	0.049554 (±0.112;0.44)	0.042564 (±0.136;0.31)	0.026043 (±0.107;0.24)
β_9 Vertical grade	0.077815 (±0.035;2.25)	0.086784 (±0.032;2.72)	0.101143 (±0.056;1.78)	0.093379 (±0.036;2.63)	-----	-----
β_{14} Vertical grade × Length of original grade	0.033973 (±0.019;1.81)	0.027904 (±0.019;1.45)	-----	-----	-----	-----
β_{11} Deviation of inside or outside shoulder width per direction from 12 ft	0.085763 (±0.045;1.90) [paved inside]	0.070920 (±0.044;1.61) [paved inside]	0.153900 (±0.070;2.20) [paved inside]	0.083181 (±0.052;1.59) [paved inside]	0.034061 (±0.020;1.72) [stabilized outside]	0.027521 (±0.016;1.68) [stabilized outside]
β_{12} Percent Trucks (e.g., 10)	-0.025233 (±0.005;-4.70)	-0.026532 (±0.005;-4.96)	-0.093899 (±0.014;-6.82)	-0.084985 (±0.010;-8.57)	-0.026276 (±0.007;-3.96)	-0.023492 (±0.005;-4.38)
Dispersion Parameter ($\hat{\delta}$)		0.94652 (±0.107;8.89)		0.58397 (±0.064;9.07)		1.21512 (±0.212;5.73)
T_{score}	13.49 > 1.96		20.12 > 1.96		8.51 > 1.96	
$L(\hat{\beta})$	-3771.0	-3246.2	-2741.9	-1485.1	-2545.8	-2398.9
AIC Value	7566.0	6518.4	5509.7	2998.3	5111.6	4819.7
Expected vs. Observed Total Truck Accident Involvements (5-year)	1,644.3 1,643.0	1,702.6 1,643.0	1,903.9 1,904.0	2,039.5 1,904.0	788.9 789.0	843.5 789.0

Notes: (1) Values in parentheses are (adjusted) standard deviation and t-statistics of the parameters above.
(2) ----- Not included in the model.
(3) 1 mile = 1.61 km, 1 ft = 0.3048 m.

Table 25. Expected reductions in truck accident involvements on an urban Interstate and freeway road section after an improvement in one geometric design element: examples based on the negative binomial model in table 24.

Length of Original Curve (mi)	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 12°				
	Reduce 1°	Reduce 2°	Reduce 3°	Reduce 4°	Reduce 5°
0.10	5.7% (±3.3%)	11.1% (±6.2%)	16.2% (±8.7%)	21.0% (±11.1%)	25.5% (±13.1%)
0.25	6.4% (±2.3%)	12.4% (±4.3%)	18.0% (±6.0%)	23.3% (±7.5%)	28.2% (±8.8%)
0.50	7.6% (±2.7%)	14.6% (±5.0%)	21.0% (±6.9%)	27.0% (±8.6%)	32.5% (±10.0%)
0.75	8.7% (±4.7%)	16.6% (±8.7%)	23.9% (±12.0%)	30.5% (±14.8%)	36.6% (±17.2%)
≥1.00	9.8% (±7.0%)	18.7% (±12.9%)	26.7% (±17.8%)	33.9% (±22.1%)	40.4% (±26.0%)

	Vertical Grade (VG): for 2% < VG <9%				
	Reduce 1%	Reduce 2%	Reduce 3%	Reduce 4%	Reduce 5%
	8.9% (±3.2%)	17.0% (±5.9%)	24.4% (±8.1%)	31.2% (±9.9%)	37.3% (±11.4%)

	Paved Inside Shoulder Width per Direction (ISH): for ISH ≤ 12 ft				
	Increase 1 ft	Increase 2 ft	Increase 3 ft	Increase 4 ft	Increase 5 ft
	8.0% (±4.8%)	15.3% (±8.9%)	22.1% (±12.5%)	28.3% (±15.5%)	34.0% (±18.2%)

Notes: Values in parentheses are one standard deviation of the expected reductions above.
1 ft = 0.3048 m; 1 mi = 1.61 km.

Table 26. Expected reductions in truck accident involvements on an urban Interstate and freeway road section after an improvement in two geometric design elements: examples based on the negative binomial model in table 24.

Length of Original Curve (LHC) = 0.10 mi					
Vertical Grade (VG): for 2% < VG < 12%	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 9°				
	Reduce 1°	Reduce 2°	Reduce 3°	Reduce 4°	Reduce 5°
Reduce 1%	14.1% (±3.7%)	19.0% (±5.7%)	23.7% (±7.8%)	28.0% (±9.8%)	32.1% (±11.6%)
Reduce 2%	21.8% (±5.6%)	26.2% (±6.4%)	30.5% (±7.8%)	34.4% (±9.3%)	38.2% (±10.8%)
Reduce 3%	28.7% (±7.5%)	32.8% (±7.6%)	36.7% (±8.3%)	40.3% (±9.3%)	43.7% (±10.4%)
Reduce 4%	35.1% (±9.1%)	38.8% (±8.9%)	42.3% (±9.1%)	45.6% (±9.7%)	48.7% (±10.4%)
Reduce 5%	40.9% (±10.4%)	44.3% (±10.0%)	47.5% (±9.9%)	50.5% (±10.1%)	53.3% (±10.5%)

Length of Original Curve (LHC) = 0.10 mi					
Paved Inside Shoulder Width per Direction (ISH): for ISH ≤ 12 ft	Horizontal Curvature (HC) in degrees/100-ft arc: for 2° ≤ HC ≤ 9°				
	Reduce 1°	Reduce 2°	Reduce 3°	Reduce 4°	Reduce 5°
Increase 1 ft	13.2% (±5.5%)	18.2% (±7.2%)	22.9% (±9.1%)	27.3% (±10.9%)	31.4% (±12.7%)
Increase 2 ft	20.2% (±8.9%)	24.7% (±9.6%)	29.0% (±10.7%)	33.1% (±11.9%)	36.9% (±13.2%)
Increase 3 ft	26.5% (±12.1%)	30.7% (±12.2%)	34.7% (±12.7%)	38.4% (±13.4%)	41.9% (±14.2%)
Increase 4 ft	32.4% (±14.9%)	36.3% (±14.6%)	39.9% (±14.7%)	43.3% (±15.0%)	46.6% (±15.4%)
Increase 5 ft	37.8% (±17.3%)	41.4% (±16.8%)	44.7% (±16.6%)	47.9% (±16.6%)	50.8% (±16.7%)

Notes: Values in parentheses are one standard deviation of the expected reductions above.
1 ft = 0.3048 m; 1 mi = 1.61 km.

6. DATA NEEDS

INTRODUCTION

The preliminary Poisson regression model reported in chapter 4 suggested that some variations in the truck accident data were unexplainable using the existing traffic and geometric design variables from the HSIS. These unexplained variations were quantified in chapter 5 by using a negative binomial regression model. The unexplained variations were found to be significant for all three roadway types (rural Interstate, urban Interstate and freeway, and rural two-lane undivided arterial) using the data from Utah. It was concluded that the preliminary Poisson model could be improved by including additional variables and by improving the accuracy of truck exposure data. This and the following chapters conclude the first phase of this study by suggesting areas and ways in which the quality and quantity of the data in HSIS can be enhanced to improve the developed preliminary models. The following discussion is based mainly on the experience working with Utah data, and focuses on three major issues:

- Which additional variables should be collected?
- What and how to improve the quality of existing data in the HSIS?
- For each roadway class, how many truck miles are required to develop truck accidents and highway geometric design relationships with a desired level of precision? (See chapter 7.)

As indicated at the outset of this report, vehicle accidents are the results of complex interactions involving many factors: the road, the traffic, the driver(s), the vehicle(s), and the environment (e.g., weather and lighting conditions). Although the focus of this study was on establishing relationships between truck accidents and highway geometric design (of the road), there are many other factors that can affect accident rates on a road section. A comprehensive list of factors was summarized in chapter 4. Ideally, all of these variables should be considered when developing an accident model. However, many of these variables (or even their proxy variables) will probably never be available at the individual road section level. Therefore, one should accept the fact that, no matter how many explanatory variables one manages to include in model development, there are always some omitted variables, especially those variables that are "qualitative-type" in nature, such as human factors.

Traditionally, this omitted variable problem has been alleviated by developing separate models for different roadway classes and vehicle types when the data permit such an analysis. In other words, vehicle accidents were first categorized by roadway class and vehicle type, and a model was then developed for each accident category. Statistically, the idea of categorizing the accidents by roadway and vehicle types in the analysis is to create a relatively homogeneous highway environment and subject (i.e., vehicle and driver) so that the accidents occurring on this specific environment involving this specific vehicle type have relatively less variations. Therefore, fewer variables are required in the model to "explain" the variations. However, as the number of accident category increases, the number of accidents available for analysis in each category decreases and, therefore, the uncertainty of the analysis results increases. In truck accident studies, the limited number of truck accidents and truck exposure data may prohibit researchers from breaking down accident analysis by both roadway class and truck types. With very few exceptions, most of the studies were unable to categorize the accidents by time of day and weather conditions because of the lack of vehicle exposure data.

ADDITIONAL VARIABLES

Except for a few data quality problems, such as missing data, the data collected in HSIS are quite complete in terms of accident and highway geometric related variables (see chapter 3). For the Utah data, the only missing highway geometric variables that might enhance the preliminary model are superelevation and roadside design conditions such as sideslope and ditch width. It is recommended that these variables be inventoried in the Utah roadlog file for all of the road sections in the future. To prioritize the data collection plan, the collection should begin with higher order highway functional classes (e.g., Interstate and freeway), and then with those road sections with sharp curves (e.g., with horizontal curvature > 2.5 degrees per 100-ft (30.48-m) arc). On the Utah accident data, it is recommended that the direction of travel at the time of accidents be collected. With this additional accident information, the geometric effects of downgrade sections on truck accidents can be distinguished from those of upgrade sections. Also, this information is likely to be required when the effects of continuous geometric design conditions are to be studied, e.g., the effects of sharp horizontal curves following long segments of straight alignment and contiguous horizontal curves.

The HSIS is, on the other hand, limited in providing detailed truck exposure data by truck type, time of day, and weather conditions. This limitation on truck exposure data makes detailed truck accident analysis unattainable, even when detailed accident data are available by truck type, time of day, and weather condition. The following is a discussion on what and how the detailed truck exposure data could be collected.

Statistically, the criterion for selecting variables is to choose the one that can "explain" a significant proportion of the variations in the truck accident data. As indicated in chapter 3, a significant percentage of these truck accidents occurred in dark (i.e., night/dawn/dusk) conditions. This percent was the highest for the combination truck accidents on rural highways: about 45 percent. Furthermore, about 15 percent to 20 percent of the combination truck accidents occurred under either snowy or rainy conditions. Based on these statistics, it is expected that the probability of a truck involved in an accident when traveling through a particular road section may be different between daytime and nighttime, and under different weather conditions. Also, based on the results of the preliminary models and other literature, it is expected that by collecting more detailed truck exposure data (or their surrogate measures) the preliminary models presented in this report can be improved to reduce the uncertainties of the estimated truck accident involvement rates. Consequently, the precision of the predicted truck accident probabilities can be increased.

One economical approach to collect additional detailed truck exposure data is through the current HPMS traffic data collection effort conducted by individual State transportation departments. The reason is that the data collection effort for AADT and vehicle classification is already in place under the current HPMS program. For example, Utah currently has machine counts at about 72 permanent stations across the State, 40 of which provide data on volume and vehicle classification counts. There are 75 temporary counting units (or "coverage counts") which are used to conduct 48-hour counts. These temporary counting units are distributed across the year and across all types of roadways. The current vehicle axle counts and vehicle classification counts on a HPMS sample road section are obtained from automatic recording machines on a 48-hour basis. These "raw" data are then adjusted for sampling variations (including axle counts, daily, day-of-week, and seasonal variations) to obtain an average daily traffic (AADT) and the percentage of vehicles by 13 vehicle classes. What is required to obtain the additional detail in the truck exposure data is to adjust the raw data on the AADT and truck percentages separately for different truck types and for time of day. Since more detailed truck exposure data are needed for individual road sections, it is recommended that additional vehicle classification stations be included to improve the precision of the estimates. A study plan will be required to properly translate the existing "raw" data, including the axle counts and the vehicle classification counts, into truck exposure data by truck type and time of day. Also, a study will be needed to determine how many and where the additional vehicle counting stations should be added to the current HPMS program so that the sampling errors can be minimized. The study will have to take into account the temporal and spatial variations of vehicle volume and vehicle classification, based on the historical data.

One potential surrogate measure for truck exposure data under different weather conditions can possibly be obtained as a function of the current overall truck exposure data and the number of rainy and snowy days in a year. Daily precipitation data are available from the National Oceanic and Atmospheric Administration's (NOAA) climatological data base. These data are readily available on computer tapes and diskettes. Our recommendation is to select 5 to 10 weather stations that have good weather data and that represent the entire State of Utah for analysis purpose. (Note that Utah currently has roughly 100 weather stations.)

In sum, at a minimum it is recommended that the following variables be collected for future studies:

- Truck exposure distribution by truck type and time of day. At a minimum, AADT and the percentages of trucks by daytime and nighttime should be collected.
- Truck exposure by weather conditions. Surrogate variables could be constructed using information such as the number of rainy and snowy days in a year.
- Superelevation and roadside design.
- The direction of travel at the time of accidents.

Should one need to investigate truck accident phenomenon on interchanges and intersections, neither the geometric nor the traffic data on interchanges and intersections are available for Utah in HSIS. Although map documentation of interchanges on Interstate highways can be obtained for analysis purpose, one needs to match accident data to the map documentation manually. In addition, Utah officials noted that some inconsistencies in the ramp accident data could exist. However, Minnesota, another HSIS State, maintains rather complete road inventory data for intersections and interchanges.

QUALITY OF EXISTING DATA

As indicated in chapter 3, several existing variables for Utah were found to have some data quality problems. For example,

- Lane width: (38 percent of the road sections are uncoded).
- Pavement conditions: service rating (87 percent of the road sections are uncoded).

- Speed limit: Posted speed limit (79 percent of the road sections are uncoded) Impact speed (50 percent coded as zero which could be a parked vehicle or an uncoded case).

- DUI: driver alcohol percent (BAC) (these data are potentially erroneous).

In addition, the roadlog, horizontal curvature, and vertical grade are not completely inventoried. In other words, for some of the road sections, the curvature and grade information is not contained in the curvature and grade files. Furthermore, road sections on local routes are currently aggregated into "zone" data based on their respective jurisdictions. Therefore, the relationships between truck accidents and road geometric design can not be established for the local road under the current Utah road inventory data.

Although the developed models for Illinois are not presented in this report, the researchers' previous experience working with data on Illinois freeway and expressway suggested that there are several problems with the data (see chapter 3 for details). First, Illinois accident data in HSIS were available annually from 1985 to 1987, but the roadlog file contained data for 1987 only. For analysis purposes, one has to assume that the road characteristics, including AADT, did not change over the 3-year period. Second, the Illinois roadlog file does not include information on lane width; instead information on surface width is included. Lane width has to be calculated based on factors such as highway functional class, median type and width, number of lanes, pavement type, allowance for parking lanes, shoulder width, and other variables. Using the Illinois freeway and expressway data, the calculated lane widths varied from 6 to 24 ft (1.83 to 7.32 m), indicating some coding errors. Third, each road section in the Illinois roadlog file is homogeneous in terms of horizontal curvature, but not necessarily in terms of vertical grade. For each road section, the approach grade and leave grade are given. However, it was possible that approach grade is uphill while leave grade is downhill, or vice versa. In addition, a road section may contain multiple vertical grades. Finally, like the Utah roadlog data, the Illinois roadlog file does not include the entire road inventory in Illinois.

A total of 100 sample police accident reports from the Utah and Illinois State Department of Transportation (50 each) was collected to perform quality checks on the accident data stored in the HSIS. No significant data quality problem was found when comparing the data in the police reports against those recorded in the HSIS accident files. However, this effort should be continued for each individual roadway class in the future to monitor the quality of the accident data. As indicated earlier, one possible accident variable that might be of interest in the accident report that was not recorded in the Utah data base is the direction of vehicle travel at the time of accidents, e.g., eastbound, westbound, etc.

Under the direction of FHWA, it can be expected that these data quality problems will be rectified and improved progressively in the future. In the next few years, FHWA may want to lay out a sampling plan for performing quality checks on highway geometric and traffic data currently available in the HSIS. This includes spot-checks of the data quality on: (1) number of lanes, (2) lane width, (3) shoulder width and type, (4) horizontal curvature, (5) vertical grade, (6) median type and width, and (7) pavement type. This sampling plan could be integrated as part of the data collection efforts recommended for collecting additional truck exposure, travel lane, median, shoulder design, and roadside design data.

7. SAMPLE SIZE REQUIREMENTS

INTRODUCTION

In this study, the relationships between truck accidents and key highway geometric design variables were established using over 5 years of data in one State--Utah. Despite the limitations in existing Utah geometric design data, e.g., lack of variations in lane width, some encouraging preliminary relationships were developed for horizontal curvature, length of curve, vertical grade, length of grade, and shoulder width. An important question that needs to be addressed next is whether the relationships established for Utah would hold in other States as well. Without examining the data from other States, this question can only be speculated. However, in view of the differences among States in weather, terrain, socioeconomic, and law enforcement conditions, as well as the differences in accident reporting practices for nonfatal accidents and in the criteria used for classifying roadways, the relationships, to some degree, are expected to vary from one State to another or, at least, from one region to another. Therefore, it would be a logical step to extend the current research by broadening the data base to include data from other geographical areas which have different weather, terrain, and socioeconomic conditions.

Now, assume that data are to be collected from other States to develop more representative truck accident-geometric design relationships for different geographical areas in the Nation. For this study, an ideal geographical area would be an area in which weather, terrain, and socioeconomic conditions are similar so that the established truck accident-geometric design relationships can be reasonably applied to any part of the area. The recommendation of this study is to collect data from each Census Division. This recommendation is based on both theoretical and practical considerations which will be discussed in this chapter. Note that there are currently 9 Census Divisions and each Division includes 3 to 9 States.

In order to collect the data from different geographical areas so that the truck accident-geometric design relationships can be established with some desired level of accuracy within each area, a data collection plan which addresses the following questions is needed:

- For each roadway class, how many truck miles (or road sections, or truck accidents) need to be collected from each geographical area so that the relationships can be established with some desired level of accuracy?
- Given the amount of truck miles that needs to be collected for each area, how many States and which States within the area the data need to be collected from in order to establish truly representative relationships?
- What is a reasonable time frame for collecting such data and what sampling strategies can be used to reduce the data collection cost?

The objective of this chapter is to estimate the sample size requirements in terms of truck miles for developing a set of "National" models which represent the truck accident-geometric design relationships for different roadway classes and for different geographical areas with some specified levels of statistical precision. The roadway classes of interest are those three classes examined in this study: (1) rural Interstate, (2) urban Interstate and freeway, and (3) rural two-lane undivided arterial. Ideally, the model should be developed as accurately as possible. In practice, due to limited resources, there is a need to strike a balance between accuracy and cost.

Because the underlying relationships between truck accidents and geometric design change over time, the relationships developed at one time may no longer be representative in later years. For example, changes in vehicle performance, socioeconomic, legislative, and law enforcement conditions over the year would change the geometric design effects on vehicle accidents even if nothing is done to the roads. For this reason, the time frame of data collection should be limited to a couple of years (say, less than 5 years). Conceptually, in order to collect a large amount of truck miles within a limited number of years and to cover different weather and terrain conditions, it is beneficial to collect the data from as many States across the country as possible. It is also desirable from a statistical point of view to collect the data from as many States as possible since this would allow us to obtain more variation in the range of values and distributions of geometric design and traffic variables. But the more States we select, the higher the initial data collection cost would be. In this chapter, the sampling strategies as to how many States, which States the data need to be collected from, and for how long the data need to be collected within each geographical area, are discussed from both theoretical and practical viewpoints. Finally, recommended geographical areas, sampling sizes within each geographical area, and sampling strategies to reduce cost are presented.

THEORETICAL DEVELOPMENT

In this section, the sample size will be determined theoretically based on the Poisson regression models described in chapter 4. For each roadway class, the sample size is estimated in terms of truck miles required for developing a statistical model with some desired level of precision for a particular geographical area. The statistical criteria which will be used to measure the precision of the model are the significance level (α) and power ($1-\gamma$) to test hypotheses about the Poisson regression parameters. (Note that α, γ are the probabilities of type I error and type II error, respectively.) The following derivation follows that of Signorini.[77]

As presented in chapter 4, the asymptotic covariance of the estimated parameters, $\hat{\underline{\beta}}$, in the Poisson regression model can be determined as $I(\hat{\underline{\beta}})^{-1}$, where $I(\hat{\underline{\beta}})$ is the Fisher information matrix with elements:

$$
\begin{aligned}
I_{jq} &= -E\left(\left. \frac{\partial^2 L(\underline{\beta})}{\partial \beta_j \, \partial \beta_q} \right|_{\beta=\hat{\beta}} \right) & (j,q=1,2,3,\ldots,k) \\
&= -E\left(\sum_{i=1}^{n} -(v_i e^{x_i \hat{\underline{\beta}}}) x_{ij} x_{iq} \right) \\
&= E\left(\sum_{i=1}^{n} w_{ijq} \right) & \left(w_{ijq} = (v_i e^{x_i \hat{\underline{\beta}}}) x_{ij} x_{iq} \right) \\
&= nE\left(\frac{1}{n} \sum_{i=1}^{n} w_{ijq} \right)
\end{aligned}
\tag{35}
$$

which is a function of $\hat{\underline{\beta}}$ and, given $\hat{\underline{\beta}}$, it does not involve dependent variable y_i. Note that n is the total number of homogeneous road sections collected from the geographical area of interest. Also, recall that the same road section in different years is considered as independent sections even if nothing has changed, and that $v_i = 365 \times AADT_i \times (T\%_i/100) \times \ell_i$.

Now, consider v_i and x_{ij}, $j=1,2,\ldots,k$, as realizations of random variables V and X_j, $j=1,2,3,\ldots,k$, which are distributed with some joint probability density or mass function. (Note that random variable V is a function of other random variables: AADT, percent trucks, and section length.) In addition, since w_{ijq} is a function of v_i and x_{ij}, where $j=1,2,\ldots,k$, it can also be regarded as a realization of a random variable, say W_{jq}, which is a function of random variables V and X_j, $j=1,2,\ldots,k$, and regression parameter $\hat{\underline{\beta}}$. Thus, I_{jq} can be reexpressed as:

$$
I_{jq} = nE(W_{jq}) \tag{36}
$$

where $E(W_{jq})$ is the expected value of random variable W_{jq}. In theory, for a given joint probability distribution of V and X_j, $j=1,2,\ldots,k$, and a regression parameter vector $\hat{\underline{\beta}}$, $E(W_{jq})$ can be assessed analytically. In practice, the analytical value of $E(W_{jq})$ is very hard to derive, unless the probability distribution of V and X_j, $j=1,2,\ldots,k$, happens to be multivariate normal, and that V is independent of X_j for $j=1,2,\ldots,k$. As will be seen later, for the highway geometric design and traffic variables considered in this study the multivariate normal distribution is not a reasonable assumption. Fortunately, given the probability distribution of V and X_j, $j=1,2,\ldots,k$, $E(W_{jq})$ can always be evaluated numerically through computer simulations. That is, by generating a fairly large number of computer-simulated values of v_i and x_{ij}, where $j=1,2,\ldots,k$, it can be approximated as:

$$E(W_{jq}) \approx \frac{1}{N}\sum_{i=1}^{N}(v_i e^{x_i\hat{\beta}})x_{ij}x_{iq} \qquad \text{as } N \to \text{large} \tag{37}$$

In this study, it was found that N=60,000 is sufficient for rural Interstate and urban Interstate and freeway, and N=120,000 is sufficient for rural two-lane undivided arterial.

For ease of exposition, let $I(\hat{\underline{\beta}}) = nF(\hat{\underline{\beta}})$, where $F(\hat{\underline{\beta}})$ is a $k \times k$ matrix with elements $E(W_{jq}), j,q=1,2,...,k$. Now, assuming that the number of road sections n is large enough to apply the asymptotic results derived above, the asymptotic variance of $\hat{\beta}_j$ is given by the jth diagonal term of $I(\hat{\underline{\beta}})^{-1}$, and, therefore, can be written as:

$$Var(\hat{\beta}_j) = n^{-1}\, d_j(\hat{\underline{\beta}}) \tag{38}$$

where $d_j(\hat{\underline{\beta}})$ is the jth diagonal term of $F(\hat{\underline{\beta}})^{-1}$.

Let x_{ij} and β_j be the explanatory variable and regression parameter of interest. Further, suppose that we wish to test the null hypothesis H_o: $\underline{\beta} = \tilde{\beta} = (\beta_1=\hat{\beta}_1, \beta_2=\hat{\beta}_2, ..., \beta_{j-1}=\hat{\beta}_{j-1}, \beta_j=0, \beta_{j+1}=\hat{\beta}_{j+1}, ..., \beta_k=\hat{\beta}_k)'$ against the alternative hypothesis H_A: $\underline{\beta} = \hat{\underline{\beta}} = (\beta_1=\hat{\beta}_1, \beta_2=\hat{\beta}_2, ..., \beta_{j-1}=\hat{\beta}_{j-1}, \beta_j=\hat{\beta}_j, \beta_{j+1}=\hat{\beta}_{j+1}, ..., \beta_k=\hat{\beta}_k)'$, at a significance level α and a power of at least 1-γ. It can be shown that this precision can be satisfied approximately for each geographical area if the number of homogeneous road sections:

$$n \geq \left[d_j(\tilde{\underline{\beta}})^{1/2} z_\alpha + d_j(\hat{\underline{\beta}})^{1/2} z_\gamma\right]^2 / \hat{\beta}_j^2 \tag{39}$$

where z_α and z_γ are respectively the 100(1-0.5α) and 100(1-0.5γ) percentiles of a standard normal distribution. Once the required number of road sections n is determined, the total truck miles needed to be collected can be computed as n times the expected truck miles of a road section, i.e., $nE(V)$, where $E(V)$ again can be evaluated numerically. If overdispersion exists as that described in chapter 4, then the required sample size must be increased by a factor of τ, where τ is the Wedderburn's overdispersion parameter and $\tau > 1$. That is, the required truck miles (TM) for each geographical area is computed as:

$$\begin{aligned} TM &= (\tau \times n)E(V) \\ &\approx (\tau \times n)\left(\frac{1}{N}\sum_{i=1}^{N}\left[365 \times AADT_i \times (T\%_i/100) \times \ell_i\right]\right) \end{aligned} \tag{40}$$

where n is the smallest integer that satisfies eq (39) and N is a large number described earlier.

It is clear from eq (39) that different explanatory variables may require different sample sizes for achieving a specific precision level. In addition, it suggests that the sample size requirement depends on (1) the desired level of precision, controlled by α and γ, (2) the range of values and distributions of explanatory variables and their cross-correlations, or, more precisely, the joint probability distribution of geometric design, traffic, and truck exposure variables, reflected in $d_j(\underline{\hat{\beta}})$ and $d_j(\underline{\hat{\beta}})$, and (3) the level of association between individual explanatory variables and truck accident involvement rate, determined by the magnitude of the regression parameter $\hat{\beta}_j$.

SAMPLE SIZE ESTIMATION

Before collecting the data and examining the truck accident-geometric design relationships in other States, the range of explanatory variable values and their distributions, and the estimates of regression parameters are not known. A reasonable approach at this point is to make sensible "guesses" of these unknowns based on the available information. For example, the ranges and distributions of explanatory variable values can be reasonably estimated using both Utah data and HPMS data (such as those published in tables HM-53 through HM-59 of *Highway Statistics*), and the regression parameters can be estimated using the values obtained in this study and other studies when applicable.[23] In what follows, the estimates of sample size requirements are derived for each of the three roadway classes, and three precision levels are considered: (1) $\alpha = 0.05$, $\gamma = 0.05$, (2) $\alpha = 0.10$, $\gamma = 0.10$, and (3) $\alpha = 0.20$, $\gamma = 0.20$.

Table 27 shows the assumed regression parameter values for the three roadway classes. For most of the explanatory variables, the estimated regression parameters of the Utah models were used. For the parameters of those geometric design variables for which the Utah data were unable to develop reasonable relationships due to lack of variations, some assumptions were made. For example:

- For rural and urban Interstate/freeways, paved outside shoulder width was assumed to have the same effects on truck accident involvements as the paved inside shoulder width.

- For rural two-lane undivided arterial, the effects of vertical grade and length of grade were assumed to be the same as those for rural Interstate.

- For rural two-lane undivided arterial, the effect of lane width was assumed to be the same as that derived in Zegeer, et al.[53] (Note that, in Zegeer, et al., the lane width effect was estimated for all vehicles, not just for trucks. Because lane width is expected to have more effect on large trucks than other vehicles, this assumption is likely to be somewhat understating the effect of lane width on truck accidents.)

There was no intention to estimate the lane width effect on truck accidents for either rural Interstate or urban Interstate and freeway because about 97 percent of these highways in the Nation have 12-ft (3.66-m) lane width (see table HM-53 in *Highway Statistics*)[23].

Tables 28 through 30 show the assumed distributions of AADT, percent trucks, section length, horizontal curvature, vertical grade, length of original curve, length of original grade, paved inside shoulder width, paved outside shoulder width, number of lanes, and lane width, that are expected to be collected from other States for each of the three roadway classes. These distributions were mainly estimated using the distributions of Utah data and the HPMS data published in *Highway Statistics*:[23]

- AADT distributions were based on the HPMS data published in table HM-57 of *Highway Statistics*.

- Percent trucks distributions were based on Utah data.

- Distributions of road section length, horizontal curvature, vertical grade, length of original curve, and length of original grade were estimated based on the Utah data. Since data will be collected from more than one State, some adjustments for the distributions of horizontal curvature and vertical grade were made to reflect the possibility of obtaining more variations in these variables.

- Distributions of paved inside shoulder width for Interstate and freeways were estimated using Utah data.

  Distributions of paved outside shoulder width for Interstate and freeways were assumed using engineering judgement.

  The distribution of outside shoulder width for rural two-lane undivided arterial was based on Utah data.

- About 92 percent of rural Interstate highways have four lanes (in table HM-55, *Highway Statistics*). Therefore, in determining sample size requirements, all rural Interstate highways were assumed to have four lanes.

  For urban Interstate and freeway, the distribution for the number of lanes was based on both the Utah data and HPMS data (in table HM-55, *Highway Statistics*).

- The distribution of lane width for rural two-lane undivided arterial was based on the HPMS data (in table HM-59, *Highway Statistics*).

In examining the Utah data, three important cross-correlations between explanatory variables were observed: (1) as AADT increases, the percent of trucks tends to decrease; (2) sharp curves were usually designed to be short in length; and (3) steep grades were typically designed to be short in length. In determining these distributions, these cross-correlations were considered to the fullest extent possible.

For road sections of the same cell in these tables, the explanatory variable values of these road sections were assumed to be uniformly distributed within the range of the cell. This assumption of a uniform distribution would usually give more conservative estimates of sample size requirements than other distributions such as truncated normal, exponential, or multinomial distributions. That is, the estimated sample size requirements are more likely to be larger than those actually required for achieving a specific precision level.

Based on the distributions of explanatory variables tabulated in tables 28 through 30, and the regression parameter values and overdispersion parameters given in table 27, the estimated sample size requirements at three levels of precision were generated using eq (40) for the three roadway classes of interest. The estimates are given in table 31. As expected, these estimates varied over explanatory variables.

By examining table 31, it is recommended that, for each geographical area where weather and socioeconomic conditions may be regarded as similar, 3.8, 2.5, and 2.4 billion truck miles (BTM) be collected for rural Interstate, urban Interstate/freeway, and rural two-lane undivided arterial, respectively. The recommended sample sizes would allow the relationships between truck accidents and most explanatory variables in each area to be developed at $\alpha = 0.10$ and $\gamma = 0.10$ with very few exceptions.

PRACTICAL CONSIDERATIONS AND RECOMMENDED SAMPLING STRATEGY

Appendix A tabulates annual truck miles by State and highway functional classes based on the 1990 HPMS data (one table for each Census Division). By examining these tables, one can observe that it is not possible to accumulate a sufficient amount of truck miles as recommended by this study in a few years if small States are selected for data collection. For example, it would require about 26 years for Vermont to accumulate 3.8 BTM on her rural Interstates. One sampling strategy is, therefore, to select large States where large amounts of truck travel could be accumulated in a few years. This would also reduce the number of States to be selected within each area and, therefore, the initial data acquisition costs could likely be reduced.

In order to collect the amount of truck miles recommended above within 2 to 3 years time frame, and at the same time to cover as much weather and socioeconomic variation as possible, our recommendation is that the data be collected for each of the 9 Census Divisions. (Note that for each Census Division one model will be developed to represent truck accident-geometric design relationships for each roadway class. And the developed model will be used by every State in the Division. If necessary, for those States for which data are not collected, model modifications such as that proposed in chapter 4 (eq 22) can be used.)

Another important sampling strategy to reduce the cost is to use the data from the existing HSIS States. This would significantly reduce the initial data acquisition costs since the data from the five HSIS States have been acquired by the FHWA and are expected to be updated, maintained, and evaluated constantly in the future.

Based on the considerations discussed above, the recommended States for data collection in each Census Division are presented in table 32. The number of years needed to collect the recommended truck miles in each Division are also presented in the table. Typically, 2 years are needed. Because most of the States in the New England Division (Census Division 1) and the Middle Atlantic Division (Census Division 2) are small States with limited highway mileage, these two Divisions were combined as one Division. Under this recommended data collection plan, 17 States, including all 5 HSIS States, were recommended for data collection (figure 3). In the East North Central Division (Census Division 3), the recommended States for data collection were Illinois and Michigan, both of which are HSIS States. If a pilot data collection is to be conducted, this is the best Division with which to start. The reason is that the road inventory data of these two States are largely available in the HSIS data base, and it is certain that their accident data can be linked with the roadlog data. The main effort will then be filling in the gaps in geometric design data. For example, as discussed in chapter 3, the current Illinois data in the HSIS do not include information on lane width, and horizontal curvature and vertical grade are recorded only for those road sections that are considered to be "potentially substandard."

ADDITIONAL REMARKS

- Overall, it is expected that the recommended sample sizes for each Census Division are somewhat conservative. The main reason is that there is evidence that Utah may have a lower truck accident involvement rate than other States. This was seen in table 9 of chapter 3. Also, by examining the HSIS data base, it was found that Utah had the lowest overall truck accident involvement rate from 1985 to 1987 among the five HSIS States. Therefore, for the amounts of truck miles recommended, on average more truck accident involvements in other States than in Utah can be observed.

- The Utah data indicated that the number of trucks involved in fatal and injury accidents is about one-third of total truck accident involvements for all three roadway classes examined (see table 8 in chapter 3). This suggests that if the focus is on fatal and injury accidents only, then the sample size requirements may have to be increased by a factor of 3 to achieve the same precision. However, further analysis with limited numbers of fatal and injury accidents from Utah rural Interstate and urban Interstate and freeway indicated that, in general, geometric design variables have a relatively more significant effect on fatal and injury truck accidents than that on property damage only (PDO) accidents. Also, it was found that by considering only fatal and injury accidents the overdispersion problem reduced quite significantly when compared to the original Poisson regression models which included PDO accidents. Overall, it was found that the sample size requirements must be increased by a factor of 1.5 to 2, instead of 3, for achieving the same precision. This means that if the focus is to be on fatal and injury accidents only, then 3 to 4 years of data in each Census Division have to be collected, instead of 2 years. Alternatively, FHWA may want to consider combining two neighboring Census Divisions into one Division and still collect the data for 2 years. This would, however, make the developed models less accurate when applied to individual States within the enlarged Division.

- In the last chapter (on Data Needs), it was recommended that truck exposure data be collected by truck type and time of day, and that roadside information be added to the data base. This recommendation should be considered in preparing the data collection plan. These additional information would increase the precision level of the developed models.

- The HSIS is currently in the process of being expanded from the current five States to between seven and nine States. Once these additional HSIS States are determined, the data collection plan presented in table 32 can be modified to include these new HSIS States.

- The initial data acquisition will be the biggest cost item. This will include the cost to acquire the accident, roadway inventory, and traffic data from the State, the cost to clean and merge these data, and the cost to fill in data gaps. Once the data base is set up, the maintenance and updating costs are expected to be relatively marginal when compared to the initial data acquisition cost. Roadway data can be updated every few years, and new accident data can to be added to the data base annually.

- When roadway data collection is conducted, it is recommended that road segments be recorded every time there is a change in any of the geometric/roadside design elements or traffic variables. This would allow the study on the effect of continuous geometric design conditions, such as sharp horizontal curves following long segments of straight alignment, contiguous horizontal curves, and grades on curves.

SUMMARY

It was recommended in this chapter that, for each Census Division, 3.8, 2.5, and 2.4 BTM be collected for rural Interstate, urban Interstate/freeway, and rural two-lane undivided arterial, respectively. The recommended sample sizes would allow the relationships between truck accidents and most of the key geometric and traffic variables to be developed at a good precision level (α = 0.10 and γ = 0.10) with very few exceptions. Based on several considerations, the recommended States for data collection in each Census Division were presented in table 32. The number of years needed to collect the recommended truck miles in each Division was also presented in the table. Typically, 2 years will be needed.

If the focus is on fatal and injury accidents only, then the sample size requirements must be increased by a factor of between 1.5 and 2. This means that 3 to 4 years of data from the

States recommended in table 32 have to be collected (instead of 2 years). Alternatively, FHWA may want to consider combining two neighboring Census Divisions into one Division and still collect the data for 2 years. This would, however, make the developed models less accurate when applied to individual States within the enlarged Division.

In the East North Central Division (Census Division 3), the recommended States for data collection were Illinois and Michigan, both of which are HSIS States. If a pilot data collection is to be conducted, this is the best Division with which to start. The reason is that the road inventory data of these two States are largely available in the HSIS data base, and it is certain that their accident data can be linked with the roadlog data. The main effort will then be filling in the gaps in geometric design data, such as those described in chapter 3.

Table 27. Parameters of the Poisson regression models used for estimating sample size requirements.

Variable	Rural Interstate	Urban Interstate & Freeway	Rural Two-Lane Undivided Arterial
Dummy intercept	-0.797996*	-1.629968*	-0.2886773*
AADT per lane (10^3 vehicles)	0.024400	0.124950	0.102226
Number of lanes	-------	0.016375	-------
Horizontal curvature (degrees/100-ft arc)	0.088861	0.039327	0.094931
Length of original curve (mi)	-------	------	-----
(Horizontal curvature) × (Length of original curve) (degrees/100-ft arc × mi)	0.234209	0.128738	0.042564
Vertical grade (percent)	0.077815	0.101143	0.077815**
Length of original grade (mi)	-------	-------	-------
(Vertical grade) × (Length of original grade) (percent × mi)	0.033973	-------	0.033973**
Deviation of paved inside shoulder width from 12 ft (ft)	0.085763	0.153900	-------
Deviation of paved or stabilized outside shoulder width from an ideal width of 12 ft (ft)	0.085763*** [paved]	0.153900*** [paved]	0.034061 [stabilized]
Percent trucks (e.g., 10)	-0.025233	-0.093899	-0.026276
Deviation of lane width from 12 ft (ft)	------- (assume 12 ft)	------- (assume 12 ft)	0.129426****
Overdispersion parameter (τ)	1.57	2.92	1.85

Notes:
- * The intercept parameters were based on the average parameter value of 1987 through 1989 dummy variables in table 24 and were adjusted if new variables from other sources were included.
- ** Assumed having the same effects as in rural Interstate highways.
- *** Paved outside shoulder width was assumed to have the same effects as the paved inside shoulder width.
- **** Adopted from Zegeer et al., equation (22), page 136.[51]

1 mi = 1.61 km, 1 ft = 0.3048 m.

Table 28. Assumed distributions of explanatory variables in percent: homogeneous rural Interstate road sections.

Percent Trucks	AADT (in 1000's of vehicles)				Total
	5-10	10-20	20-30	30-40	
10-20	5	10	15	5	35
20-30	10	20	5	5	40
30-40	10	10	5	0	25
Total	25	40	25	10	100

Horizontal Curvature (HC) (degrees/100-ft arc)	Road Section Length (in miles)						Total
	0.01-0.05	0.05-0.10	0.10-0.25	0.25-0.50	0.50-1.0	1.0-1.5	
HC≤1.0	10	10	15	15	15	5	70
1.0<HC≤2.5	5	0	0	5	0	5	15
2.5<HC≤5.5	5	0	5	0	0	0	10
5.5<HC≤8.5	0	5	0	0	0	0	5
Total	20	15	20	20	15	10	100

Vertical Grade (VG) (percent)	Road Section Length (in miles)						Total
	0.01-0.05	0.05-0.10	0.10-0.25	0.25-0.50	0.50-1.0	1.0-1.5	
VG≤2.5	10	10	15	15	15	10	75
2.5<VG≤4.5	5	0	5	5	0	0	15
4.5<VG≤6.5	0	5	0	0	0	0	5
6.5<VG≤8.5	5	0	0	0	0	0	5
Total	20	15	20	20	15	10	100

Paved Shoulder Width (SW) (ft)	Inside (Left or Median)	Outside (Right)
SW≤3	5	0
3<SW≤5	80	0
5<SW≤7	10	20
7<SW≤9	5	20
9<SW≤11	0	60
Total	100	100

Notes: (1) Number of lanes was assumed to be four. (2) Length of original curve was assumed to be three times that of road section length. (3) Length of original grade was assumed to be four times that of road section length. (4) 1 mi = 1.61 km, 1 ft = 0.3048 m.

Table 29. Assumed distributions of explanatory variables in percent: homogeneous urban Interstate and freeway road sections.

Percent Trucks	AADT (in 1000's of vehicles)					Total
	10-15	15-35	35-60	60-100	100-120	
1-5	0	5	0	5	10	20
5-10	0	5	10	10	5	30
10-20	5	15	15	5	0	40
20-30	5	5	0	0	0	10
Total	10	30	25	20	15	100

Horizontal Curvature (HC) (degrees/100-ft arc)	Road Section Length (in miles)						Total
	0.01-0.05	0.05-0.10	0.10-0.25	0.25-0.50	0.50-1.0	1.0-1.5	
HC≤1.0	5	15	25	15	5	5	70
1.0<HC≤2.5	0	5	5	0	5	0	15
2.5<HC≤5.5	5	0	5	0	0	0	10
5.5<HC≤8.5	5	0	0	0	0	0	5
Total	15	20	35	15	10	5	100

Vertical Grade (VG) (percent)	Road Section Length (in miles)						Total
	0.01-0.05	0.05-0.10	0.10-0.25	0.25-0.50	0.50-1.0	1.0-1.5	
VG≤2.5	10	15	25	10	5	5	70
2.5<VG≤4.5	5	0	5	0	5	0	15
4.5<VG≤6.5	0	0	5	5	0	0	10
6.5<VG≤8.5	0	5	0	0	0	0	5
Total	15	20	35	15	10	5	100

Paved Shoulder Width (SW) (ft)	Inside (Left or Median)	Outside (Right)
SW≤3	5	0
3<SW≤5	80	0
5<SW≤7	10	20
7<SW≤9	5	20
9<SW≤11	0	60
Total	100	100

No. of Lanes	AADT (in 1000's of vehicles)					Total
	10-15	15-35	35-60	60-100	100-120	
4	10	20	10	5	0	45
6	0	10	15	15	10	50
8	0	0	0	0	5	5
Total	10	30	25	20	15	100

Notes: (1) Length of original curve was assumed to be three times that of road section length. (2) Length of original grade was assumed to be four times that of road section length. (3) 1 mi = 1.61 km, 1 ft = 0.3048 m.

Table 30. Assumed distributions of explanatory variables in percent: homogeneous rural two-lane undivided arterial sections.

Percent Trucks	AADT (in 1000's of vehicles) 1-2	2-3	3-10	10-15	Total
5-10	5	5	15	5	30
10-20	15	5	10	0	30
20-30	10	10	5	0	25
30-40	10	0	5	0	15
Total	40	20	35	5	100

Horizontal Curvature (HC) (degrees/100-ft arc)	Road Section Length (in miles) 0.01-0.05	0.05-0.10	0.10-0.25	0.25-0.50	0.50-1.0	1.0-1.5	Total
HC≤2.5	10	20	25	5	5	5	70
2.5<HC≤5.5	5	0	5	0	5	0	15
5.5<HC≤8.5	0	5	0	5	0	0	10
8.5<HC≤15.5	5	0	0	0	0	0	5
Total	20	25	30	10	10	5	100

Vertical Grade (VG) (percent)	Road Section Length (in miles) 0.01-0.05	0.05-0.10	0.10-0.25	0.25-0.50	0.50-1.0	1.0-1.5	Total
VG≤2.5	10	15	20	10	5	5	65
2.5<VG≤4.5	5	0	5	0	5	0	15
4.5<VG≤6.5	0	5	5	0	0	0	10
6.5<VG≤8.5	5	5	0	0	0	0	10
Total	20	25	30	10	10	5	100

Stabilized Shoulder Width (SW) (ft)	Outside (Right)
SW≤3	45
3<SW≤5	15
5<SW≤7	20
7<SW≤9	15
9<SW≤11	5
Total	100

Lane Width (ft)	Percent
10	10
11	15
12	70
13	5
Total	100

Notes: (1) Length of original curve was assumed to be three times that of road section length. (2) Length of original grade was assumed to be four times that of road section length. (3) 1 mi = 1.61 km, 1 ft = 0.3048 m.

Table 31. Required sample size in terms of billion truck miles (BTM) for achieving some specified precision levels.

Variable	Rural Interstate			Urban Interstate & Freeway			Rural Two-lane Undivided Arterial		
	α=0.05 γ=0.05	α=0.10 γ=0.10	α=0.20 γ=0.20	α=0.05 γ=0.05	α=0.10 γ=0.10	α=0.20 γ=0.20	α=0.05 γ=0.05	α=0.10 γ=0.10	α=0.20 γ=0.20
AADT per lane	8.7	6.1	3.7	0.7	0.5	0.3	0.7	0.5	0.4
Number of lanes	----	----	----	2.3	1.6	1.0	----	----	----
Horizontal curvature	5.4	3.8	2.3	----	----	----	0.6	0.5	0.3
Horizontal curvature × Length of curve	1.4	1.0	0.6	5.7	4.0	2.5	3.4	2.4	1.5
Vertical grade	6.1	4.3	2.6	0.8	0.6	0.4	1.7	1.2	0.8
Vertical grade × Length of grade	7.9	5.6	3.4	----	----	----	3.4	2.4	1.5
Paved inside shoulder width	2.7	1.9	1.2	1.1	0.8	0.5	----	----	----
Paved outside shoulder width	1.2	0.9	0.5	0.4	0.3	0.2	2.3	1.7	1.0
Percent trucks	0.5	0.4	0.2	0.1	0.1	0.1	0.3	0.2	0.2
Lane width	----	----	----	----	----	----	2.2	1.6	1.0

Notes: Shaded values are the recommended truck miles to be collected for the roadway class above.
---- Very large sample size is required.

Table 32. Recommended States in each Census Division and number of years necessary to collect the data.

Census Division	Recommended States	Annual Truck Miles[3] (Million Truck Miles)	No. of Years Needed	An HSIS State[2]
1. New England[1] (ME,VT,RI,NH,MA, CT)	ME	Rural Interstate = 230 Urban Interstate & Freeway = 40 Rural 2-Ln Undivided Arterial = 240	2 2 2	Yes
2. Middle Atlantic[1] (NY,NJ,PA)	PA	Rural Interstate = 2,260 Urban Interstate & Freeway = 1,440 Rural 2-Ln Undivided Arterial = 1,390	2 2 2	No
3. East North Central (OH,IN,IL,MI,WI)	IL	Rural Interstate = 1,930 Urban Interstate & Freeway = 1,500 Rural 2-Ln Undivided Arterial = 730	2 1 2	Yes
	MI	Rural Interstate = 730 Urban Interstate & Freeway = 1,230 Rural 2-Ln Undivided Arterial = 700	2 1 2	Yes
4. West North Central (MN,IA,MO,ND,SD, NE,KS)	MN	Rural Interstate = 400 Urban Interstate & Freeway = 450 Rural 2-Ln Undivided Arterial = 770	3 2 2	Yes
	MO	Rural Interstate = 1,350 Urban Interstate & Freeway = 740 Rural 2-Ln Undivided Arterial = 930	2 2 2	No
5. South Atlantic (DE,MD,DC,VA, WV,NC,SC,GA, FL)	FL	Rural Interstate = 1,760 Urban Interstate & Freeway = 1,170 Rural 2-Ln Undivided Arterial = 1,390	2 1 1	No
	NC	Rural Interstate = 1,400 Urban Interstate & Freeway = 870 Rural 2-Ln Undivided Arterial = 960	2 1 1	No
6. East South Central (KY,TN,AL,MS)	KY	Rural Interstate = 1,170 Urban Interstate & Freeway = 410 Rural 2-Ln Undivided Arterial = 530	2 2 2	No
	TN	Rural Interstate = 1680 Urban Interstate & Freeway = 700 Rural 2-Ln Undivided Arterial = 650	2 2 2	No
7. West South Central (AR,LA,OK,TX)	LA	Rural Interstate = 950 Urban Interstate & Freeway = 540 Rural 2-Ln Undivided Arterial = 740	2 1 1	No
	TX	Rural Interstate = 2,300 Urban Interstate & Freeway = 2,700 Rural 2-Ln Undivided Arterial = 2,730	2 1 1	No
8. Mountain (MT,ID,WY,CO,NM, AZ,UT,NV)	AZ	Rural Interstate = 1,160 Urban Interstate & Freeway = 420 Rural 2-Ln Undivided Arterial = 340	2 2 2	No
	UT	Rural Interstate = 580 Urban Interstate & Freeway = 270 Rural 2-Ln Undivided Arterial = 210	2 3 3	Yes
	WY	Rural Interstate = 630 Urban Interstate & Freeway = 30 Rural 2-Ln Undivided Arterial = 280	2 2 2	No
9. Pacific (WA,OR,CA,AK,HI)	CA	Rural Interstate = 2,250 Urban Interstate & Freeway = 5,090 Rural 2-Ln Undivided Arterial = 2,730	2 1 1	No
	WA	Rural Interstate = 460 Urban Interstate & Freeway = 790 Rural 2-Ln Undivided Arterial = 410	2 1 1	No

[1] Census Divisions 1 and 2 are to be combined as one Division.
[2] As of September 1992.
[3] Estimates from 1990 HPMS statewide data provided by FHWA.

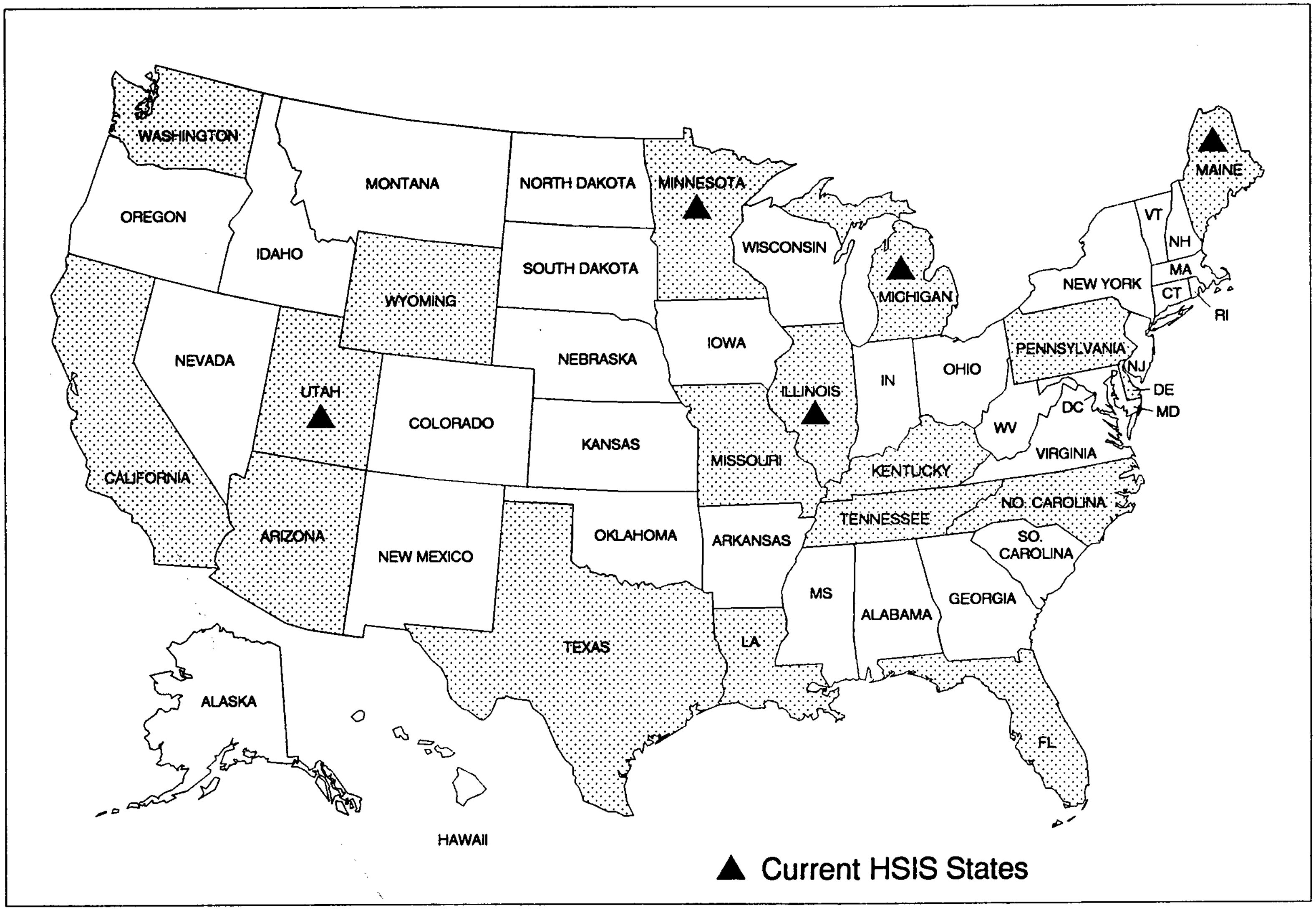

Figure 3. States recommended for data collection.

APPENDIX A. ANNUAL TRUCK TRAVEL BY STATE AND BY HIGHWAY FUNCTIONAL CLASS

Table 33. Census Division: New England.
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
CONNECTICUT	Rural Interstate	72.2	114.7	186.9
	Rural Other Principal Arterial	30.5	22.6	53.1
	Rural Minor Arterial	29.7	29.7	59.4
	Urban Interstate	243.0	447.7	690.8
	Urban Other F&E	101.6	48.4	149.9
MAINE	Rural Interstate	76.9	152.0	228.9
	Rural Other Principal Arterial	80.3	66.4	146.7
	Rural Minor Arterial	86.4	51.5	137.9
	Urban Interstate	15.3	14.4	29.6
	Urban Other F&E	4.2	5.6	9.8
MASSACHUSETTS	Rural Interstate	66.7	243.8	310.5
	Rural Other Principal Arterial	35.4	28.0	63.3
	Rural Minor Arterial	31.4	29.5	60.9
	Urban Interstate	238.0	496.7	734.6
	Urban Other F&E	80.3	93.7	174.0
NEW HAMPSHIRE	Rural Interstate	42.1	51.6	93.8
	Rural Other Principal Arterial	39.9	35.6	75.5
	Rural Minor Arterial	59.4	46.0	105.4
	Urban Interstate	20.2	51.8	72.0
	Urban Other F&E	35.6	20.3	55.9
RHODE ISLAND	Rural Interstate	13.4	8.2	21.6
	Rural Other Principal Arterial	7.4	6.8	14.3
	Rural Minor Arterial	15.4	3.5	18.9
	Urban Interstate	41.0	13.7	54.7
	Urban Other F&E	27.7	61.7	89.4
VERMONT	Rural Interstate	45.8	100.0	145.4
	Rural Other Principal Arterial	32.4	21.8	54.3
	Rural Minor Arterial	40.0	26.4	66.5
	Urban Interstate	10.5	7.6	18.1
	Urban Other F&E	1.7	0.5	2.2

Table 34. Census Division: Middle Atlantic.
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
NEW JERSEY	Rural Interstate	118.5	187.4	305.9
	Rural Other Principal Arterial	85.7	99.3	185.0
	Rural Minor Arterial	88.6	69.6	158.2
	Urban Interstate	258.4	266.8	525.2
	Urban Other F&E	138.4	143.7	282.1
NEW YORK	Rural Interstate	163.5	823.3	986.8
	Rural Other Principal Arterial	137.0	347.0	484.0
	Rural Minor Arterial	224.6	238.7	463.3
	Urban Interstate	445.2	667.8	1113.0
	Urban Other F&E	360.3	619.7	980.0
PENNSYLVANIA	Rural Interstate	398.4	1864.3	2262.6
	Rural Other Principal Arterial	359.8	487.2	847.0
	Rural Minor Arterial	489.5	265.6	755.1
	Urban Interstate	319.9	796.1	1116.0
	Urban Other F&E	158.1	163.6	321.7

Table 35. Census Division: East North Central.
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
ILLINOIS	Rural Interstate	171.8	1755.0	1926.7
	Rural Other Principal Arterial	141.9	315.7	457.6
	Rural Minor Arterial	167.4	214.4	381.8
	Urban Interstate	479.9	959.8	1439.8
	Urban Other Free & Expressway	25.4	33.8	59.2
INDIANA	Rural Interstate	239.0	1626.8	1865.8
	Rural Other Principal Arterial	148.5	668.1	816.6
	Rural Minor Arterial	201.2	626.1	827.3
	Urban Interstate	190.0	702.1	892.2
	Urban Other Free & Expressway	25.8	80.3	106.1
MICHIGAN	Rural Interstate	137.5	592.0	729.5
	Rural Other Principal Arterial	126.4	358.9	485.3
	Rural Minor Arterial	102.4	218.8	321.2
	Urban Interstate	221.2	696.8	918.0
	Urban Other Free & Expressway	49.0	259.3	308.3
OHIO	Rural Interstate	209.1	1391.6	1600.8
	Rural Other Principal Arterial	184.7	888.6	1073.3
	Rural Minor Arterial	217.3	308.8	526.1
	Urban Interstate	279.6	698.9	978.5
	Urban Other Free & Expressway	47.0	70.5	117.5
WISCONSIN	Rural Interstate	109.3	897.9	1007.2
	Rural Other Principal Arterial	298.1	526.4	824.5
	Rural Minor Arterial	195.8	291.3	487.2
	Urban Interstate	75.4	135.2	210.6
	Urban Other Free & Expressway	52.3	93.9	146.2

Table 36. Census Division: West North Central.
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
IOWA	Rural Interstate	121.3	790.0	911.3
	Rural Other Principal Arterial	165.8	358.6	524.4
	Rural Minor Arterial	149.0	184.7	333.6
	Urban Interstate	40.3	174.6	214.9
	Urban Other Free & Expressway	0	0	0
KANSAS	Rural Interstate	71.7	415.3	487.0
	Rural Other Principal Arterial	116.0	351.6	467.6
	Rural Minor Arterial	74.8	180.0	254.8
	Urban Interstate	62.0	113.6	175.6
	Urban Other Free & Expressway	11.4	8.2	19.7
MINNESOTA	Rural Interstate	75.9	327.2	403.1
	Rural Other Principal Arterial	179.0	416.0	595.0
	Rural Minor Arterial	138.8	150.7	289.4
	Urban Interstate	146.7	184.5	331.3
	Urban Other Free & Expressway	52.3	61.0	113.3
MISSOURI	Rural Interstate	144.6	1202.8	1347.4
	Rural Other Principal Arterial	176.3	579.6	755.9
	Rural Minor Arterial	120.0	192.4	312.4
	Urban Interstate	152.6	473.5	626.1
	Urban Other Free & Expressway	48.5	68.5	117.0
NEBRASKA	Rural Interstate	68.6	459.1	527.7
	Rural Other Principal Arterial	83.2	193.4	276.6
	Rural Minor Arterial	88.3	128.7	217.0
	Urban Interstate	18.1	39.9	57.9
	Urban Other Free & Expressway	1.8	2.8	4.5
N. DAKOTA	Rural Interstate	38.8	125.5	164.3
	Rural Other Principal Arterial	35.6	63.4	99.0
	Rural Minor Arterial	56.5	69.0	125.5
	Urban Interstate	4.4	13.1	17.5
	Urban Other Free & Expressway	0	0	0
S. DAKOTA	Rural Interstate	41.5	172.9	214.4
	Rural Other Principal Arterial	55.6	126.2	181.8
	Rural Minor Arterial	40.8	58.4	99.2
	Urban Interstate	2.7	2.3	5.0
	Urban Other Free & Expressway	0	0	0

Table 37. Census Division: South Atlantic.
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
DELAWARE	Rural Interstate	0	0	0
	Rural Other Principal Arterial	49.0	82.4	131.5
	Rural Minor Arterial	22.1	36.5	58.4
	Urban Interstate	28.8	123.4	152.2
	Urban Other Free & Expressway	0	0	0
DIST. OF COL.	Rural Interstate	0	0	0
	Rural Other Principal Arterial	0	0	0
	Rural Minor Arterial	0	0	0
	Urban Interstate	6.0	1.4	7.4
	Urban Other Free & Expressway	6.6	1.3	7.9
FLORIDA	Rural Interstate	311.7	1451.8	1763.6
	Rural Other Principal Arterial	333.3	695.1	1028.4
	Rural Minor Arterial	205.7	358.6	564.3
	Urban Interstate	290.9	623.5	914.4
	Urban Other Free & Expressway	178.1	78.7	256.9
GEORGIA	Rural Interstate	317.1	1664.9	1982.0
	Rural Other Principal Arterial	169.2	374.3	543.6
	Rural Minor Arterial	278.0	513.2	791.2
	Urban Interstate	323.8	775.1	1098.9
	Urban Other Free & Expressway	55.7	66.8	122.5
MARYLAND	Rural Interstate	109.0	237.6	346.6
	Rural Other Principal Arterial	150.6	140.1	290.7
	Rural Minor Arterial	156.8	69.0	225.7
	Urban Interstate	492.2	607.1	1099.3
	Urban Other Free & Expressway	121.2	95.8	217.1
N. CAROLINA	Rural Interstate	300.9	1094.9	1395.9
	Rural Other Principal Arterial	270.3	494.6	764.9
	Rural Minor Arterial	206.3	133.1	339.4
	Urban Interstate	207.6	433.8	641.5
	Urban Other Free & Expressway	108.7	115.7	224.4

Table 37. Census Division: South Atlantic (continued).
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
S. CAROLINA	Rural Interstate	213.0	1053.3	1266.3
	Rural Other Principal Arterial	124.3	217.5	341.8
	Rural Minor Arterial	182.1	140.1	322.2
	Urban Interstate	63.2	185.2	248.4
	Urban Other Free & Expressway	6.0	4.2	10.2
VIRGINIA	Rural Interstate	228.8	1031.6	1260.4
	Rural Other Principal Arterial	223.5	529.7	753.2
	Rural Minor Arterial	235.3	126.8	362.1
	Urban Interstate	335.8	425.1	760.9
	Urban Other Free & Expressway	49.5	34.3	83.7
W. VIRGINIA	Rural Interstate	89.4	508.5	597.9
	Rural Other Principal Arterial	77.3	107.6	184.8
	Rural Minor Arterial	133.1	81.7	214.8
	Urban Interstate	36.7	113.3	150.0
	Urban Other Free & Expressway	15.2	16.0	31.2

Table 38. Census Division: East South Central.
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
ALABAMA	Rural Interstate	186.6	759.9	946.6
	Rural Other Principal Arterial	225.4	547.3	772.6
	Rural Minor Arterial	169.5	221.6	391.1
	Urban Interstate	135.6	342.3	478.0
	Urban Other Free & Expressway	9.4	8.0	17.4
KENTUCKY	Rural Interstate	197.2	972.4	1169.7
	Rural Other Principal Arterial	175.3	266.7	442.0
	Rural Minor Arterial	103.4	67.4	170.8
	Urban Interstate	117.3	245.0	362.3
	Urban Other Free & Expressway	20.0	23.3	43.3
MISSISSIPPI	Rural Interstate	64.0	567.6	631.6
	Rural Other Principal Arterial	78.2	442.2	520.4
	Rural Minor Arterial	84.4	253.1	337.5
	Urban Interstate	26.6	238.2	264.8
	Urban Other Free & Expressway	1.8	3.5	5.3
TENNESSEE	Rural Interstate	228.9	1453.9	1682.8
	Rural Other Principal Arterial	87.9	189.5	277.4
	Rural Minor Arterial	167.4	297.6	465.0
	Urban Interstate	164.2	538.0	702.2
	Urban Other Free & Expressway	0	0	0

Table 39. Census Division: West South Central.
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
ARKANSAS	Rural Interstate	75.6	779.7	855.4
	Rural Other Principal Arterial	92.4	360.2	452.6
	Rural Minor Arterial	85.4	224.4	309.8
	Urban Interstate	45.3	165.1	210.6
	Urban Other Free & Expressway	14.7	36.7	51.3
LOUISIANA	Rural Interstate	271.4	674.1	945.4
	Rural Other Principal Arterial	167.2	229.1	396.3
	Rural Minor Arterial	198.1	255.4	453.5
	Urban Interstate	170.2	325.3	495.5
	Urban Other Free & Expressway	21.6	22.8	44.4
OKLAHOMA	Rural Interstate	N/A	N/A	N/A
	Rural Other Principal Arterial	N/A	N/A	N/A
	Rural Minor Arterial	N/A	N/A	N/A
	Urban Interstate	N/A	N/A	N/A
	Urban Other Free & Expressway	N/A	N/A	N/A
TEXAS	Rural Interstate	383.1	1920.3	2303.4
	Rural Other Principal Arterial	654.9	1681.5	2336.4
	Rural Minor Arterial	293.6	507.8	801.4
	Urban Interstate	570.0	1527.6	2097.5
	Urban Other Free & Expressway	303.9	302.5	606.4

Table 40. Census Division: Mountain.
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
ARIZONA	Rural Interstate	296.6	868.1	1164.7
	Rural Other Principal Arterial	83.0	104.0	187.0
	Rural Minor Arterial	129.1	72.7	201.8
	Urban Interstate	131.6	260.5	392.1
	Urban Other Free & Expressway	28.6	1.2	29.9
COLORADO	Rural Interstate	167.9	493.2	661.1
	Rural Other Principal Arterial	124.3	278.3	402.6
	Rural Minor Arterial	89.5	131.0	220.5
	Urban Interstate	131.9	152.8	284.7
	Urban Other Free & Expressway	56.8	44.0	100.8
IDAHO	Rural Interstate	34.1	270.1	304.2
	Rural Other Principal Arterial	32.5	88.7	121.3
	Rural Minor Arterial	13.9	17.4	31.4
	Urban Interstate	12.4	51.3	63.7
	Urban Other Free & Expressway	0	0	0
MONTANA	Rural Interstate	63.6	237.9	301.5
	Rural Other Principal Arterial	72.6	84.5	157.1
	Rural Minor Arterial	106.2	74.4	180.5
	Urban Interstate	10.7	17.3	28.1
	Urban Other Free & Expressway	0	0	0
NEVADA	Rural Interstate	76.6	382.9	459.5
	Rural Other Principal Arterial	54.5	59.5	114.0
	Rural Minor Arterial	44.3	49.0	93.2
	Urban Interstate	31.5	42.1	73.6
	Urban Other Free & Expressway	21.5	15.5	37.0
NEW MEXICO	Rural Interstate	104.0	616.9	720.9
	Rural Other Principal Arterial	63.7	138.9	202.6
	Rural Minor Arterial	140.8	255.3	396.1
	Urban Interstate	42.7	291.6	334.3
	Urban Other Free & Expressway	0	0	0

Table 40. Census Division: Mountain (continued).
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
UTAH	Rural Interstate	124.2	460.0	584.2
	Rural Other Principal Arterial	45.6	117.7	163.2
	Rural Minor Arterial	33.7	46.0	79.6
	Urban Interstate	94.4	156.4	250.7
	Urban Other Free & Expressway	6.2	8.7	14.9
WYOMING	Rural Interstate	30.1	604.0	634.1
	Rural Other Principal Arterial	11.6	152.8	164.4
	Rural Minor Arterial	46.0	105.3	151.3
	Urban Interstate	4.6	22.6	27.2
	Urban Other Free & Expressway	0.4	0.1	0.5

Table 41. Census Division: Pacific.
(in Million Truck Miles)

State	Functional Class	Single-Unit	Combination	Total VMT
ALASKA	Rural Interstate	29.0	34.5	63.5
	Rural Other Principal Arterial	0.5	0.3	0.8
	Rural Minor Arterial	27.6	27.1	54.7
	Urban Interstate	12.6	8.8	21.4
	Urban Other Free & Expressway	0	0	0
CALIFORNIA	Rural Interstate	597.1	1654.3	2251.3
	Rural Other Principal Arterial	498.0	1842.6	2340.6
	Rural Minor Arterial	225.9	567.4	793.3
	Urban Interstate	666.2	2496.1	3162.4
	Urban Other Free & Expressway	623.1	1302.0	1925.1
HAWAII	Rural Interstate	3.2	2.0	5.2
	Rural Other Principal Arterial	4.5	2.2	6.7
	Rural Minor Arterial	29.1	9.7	38.8
	Urban Interstate	19.6	2.8	22.4
	Urban Other Free & Expressway	17.6	5.6	23.1
OREGON	Rural Interstate	137.2	664.4	801.6
	Rural Other Principal Arterial	195.7	364.0	559.7
	Rural Minor Arterial	85.3	95.7	181.0
	Urban Interstate	82.0	161.2	243.1
	Urban Other Free & Expressway	32.3	35.2	67.5
WASHINGTON	Rural Interstate	84.4	377.8	462.2
	Rural Other Principal Arterial	132.7	222.5	355.2
	Rural Minor Arterial	60.7	59.2	119.8
	Urban Interstate	195.4	450.8	646.3
	Urban Other Free & Expressway	57.9	84.3	142.2

REFERENCES

1. Transportation Research Board, *Providing Access for Large Trucks*, Special Report 223, Transportation Research Council, Washington, DC, 1989, 316 pp.

2. D.W. Harwood, J.G. Viner, and E.R. Russell, "A Truck Accident Rate Model for Hazardous Materials Routing," paper presented at the *Annual Meeting of the Transportation Research Board*, January 1990.

3. F.M. Council, D.W. Reinfurt, B.J. Campbell, F.L. Roediger, C.L. Carroll, A.K. Dutt, and J.R. Dunham, *Accident Research Manual*, FHWA-RD-80-016, Federal Highway Administration, Washington, DC, 1980.

4. P.S. Carroll, *Symposium on Driving Exposure*, NTIS PB-204 103, Highway Safety Research Institute, Ann Arbor, MI, 1975.

5. F.M. Council, J.R. Stewart, D.W. Reinfurt, W.W. Hunter, *Exposure Measures for Evaluating Highway Safety Issues*: Volume 1 (PR123) and Volume 2 (PR124)," Highway Safety Research Center, University of North Carolina, Chapel Hill, NC, 1983.

6. B.L. Bowman and J.E. Hummer, *Examination of Truck Accidents on Urban Freeways*, FHWA-RD-89-201, Federal Highway Administration, Washington, DC, 1989.

7. S. Khasnabis and Ramiz-Al Assar, "Analysis of Heavy Truck Accident data - Exposure Based Approach," *ASCE Journal of Transportation Engineering*, Vol. 115, No. 3, 1989, pp 299-304.

8. J.D. Thorpe, "Calculating Relative Involvement Rates in Accidents Without Determining Exposure," *Australian Road Research*, Vol. 4, No. 1, 1964, pp 25-36.

9. F.A. Haight, "Accrued Framework for Bypassing Exposure," *Journal of Safety Research*, Volume II, 1970.

10. M.J. Koornstra, "A Model for Estimation of Collective Exposure and Proneness from Accident Data," *Accident Analysis and Prevention*, Vol. 5, Pergamon Press, Great Britain, 1973.

11. P. Mengert, *Analysis and Testing of Koornstra-Type Induced Exposure Models*, Final Report FR-45-U-FAA-86-01, U.S. Department of Transportation, Transportation Systems Center, Cambridge, MA, October 1985.

12. B. Carr, "A Statistical Analysis of Rural Ontario Traffic Accidents Using Induced Exposure Data," *Accident Analysis and Prevention*, Vol. 1, Pergamon Press, Great Britain, 1969.

13. W.K. Hall, *An Empirical Analysis of Accident Data Using Induced Exposure*, HIT Lab Report, September 1970.

14. E.C. Cerrelli, "Driver Exposure - The Indirect Approach for Obtaining Relative Measures," *Accident Analysis and Prevention*, Pergamon Press, Great Britain, 1973.

15. T.L. Maleck and J.E. Hummer, "Driver Age and Highway Safety," *Transportation Research Record* 1059, Transportation Research Board, Washington, DC, 1986.

16. J.P. Eicher, H.D. Robertson, and G.R. Toth, *Large Truck Accidents Causation*, DOT-HS-806-300, National Center for Statistics and Analysis, NHTSA, U.S. Department of Transportation, Washington DC, 1982.

17. K.L. Campbell, D.F. Blower, R.G. Gattis, and A.C. Wolfe, *Analysis of Accident Rates of Heavy-Duty Vehicles*, The University of Michigan Transportation Research Institute, DTNH22-83-C-07188, Ann Arbor, MI, 1988, 123 pp.

18. National Highway Traffic Safety Administration, *Fatal Accident Reporting System: A Review of Information on Fatal Traffic Accidents in the US*, U.S. Department of Transportation, Washington, DC, 1975-1985.

19. National Highway Traffic Safety Administration, *National Accident Sampling System*, , U.S. Department of Transportation, 1980-1986 (annual).

20. Michigan Department of Transportation, *A Background Report to the State Transportation Commission on Truck Safety, Revenue and Taxation, and Truck Services and Highway Facilities*, Lansing, MI, 1986.

21. W.R. MaCasland and R.W. Stokes, *Truck Operations and Regulations on Urban Freeways*, FHWA/TX-85/28, August 1984.

22. T.F. Golob, *Physical Analysis for the Characteristics of Truck Involved Freeway Accidents - 1983-1984 in Los Angeles, Orange, and Ventura Counties*, TM-TAS-1 and TM-TAS-2, Institute of Transportation Studies, University of California, Irvine, CA, 1986.

23. Federal Highway Administration, *Highway Statistics*, U.S. Department of Transportation, 1970-1989 (annual).

24. P. Jovanis and J. Delleur, "Exposure-Based Analysis of Motor Vehicle Accidents," *Transportation Research Record* 910, Transportation Research Board, Washington, DC, 1983.

25. Transportation Research Board, *Twin Trailer Trucks*, Special Report 211, Transportation Research Council, Washington, DC, 1986, 388 pp.

26. W.S. Meyers, "Comparison of Truck and Passenger Car Accident Rates on Limited Access Facilities," *Transportation Research Record* 808, Transportation Research Board, Washington, DC, 1981.

27. D.F. Preusser and H.S. Stein, *Comparison of Passenger Vehicle and Truck Crash Rates on Toll Roads*, Dunlap and Associates, Inc., Norwolk, CT, 1987.

28. American Automobile Association, *Cars and Trucks Sharing the Roads Safely*, Dearborn, MI, 1986.

29. K.L. Campbell, O. Carsten, and R.H. Shultz, *Truck Accident Trends in Michigan, 1978-1984*, UMTRI-86-16, University of Michigan Transportation Research Institute, Ann Arbor, MI, 1986.

30. H.S. Stein and I. S. Jones, "Crash Involvement of Large Trucks by Configuration: A Case-Control Study," *American Journal of Public Health*, Vol. 78, No. 5, 1988, pp 491-498.

31. Transportation Research Board, *Truck Weight Limits*, Special Report 225, Transportation Research Council, Washington, DC, 1990, 307 pp.

32. P.P. Jovanis, H. Chang, and I. Zabaneh, "A Comparison of Accident Rates for Two Truck Configurations," paper presented at 68th *Annual Meeting of Transportation Research Board*, Washington, DC, January 1989, 29 pp.

33. V.D. Graf and K. Archuleta, *Truck Accidents by Classification*, FHWA/CA/TE-85, CALTRANS, Sacramento, and Federal Highway Administration, January 1985.

34. J.C. Glennon, *Matched Pair Analysis, Consolidated Freightways vs. Larson, 82-1230*, U.S. District Court, Middle District of Pennsylvania, 1981.

35. T. Chira-Chavala, D.E. Cleveland, and L.P. Kostyniuk, "Severity of Large-Truck and Combination-Vehicle Accidents in Over-the-Road Service: A Discrete Multivariate Analysis," *Transportation Research Record* 975, Transportation Research Board, National Research Council, Washington, DC, 1984, pp 23-26.

36. C.S. Yoo, M.L. Reiss, and H.W. McGee, *Comparison of California Accident Rates for Single and Double Tractor-Trailer Combination Trucks*, FHWA-RD-78-94, Federal Highway Administration, Washington, DC, March 1978, 70 pp.

37. Transportation Research Institute, *Proceedings of the National Truck Safety Symposium*, Presentations on the Large Truck Survey Program at the University of Michigan, 1987.

38. J. Hedlund, *The Severity of Large Truck Accidents*, DOT HS-802-332, NHTSA Technical Note, U. S. Department of Transportation, Washington, DC, 1977, 34 pp.

39. Roy Jorgensen Associates, Inc., *Cost and Safety Effectiveness of Highway Design Elements*, National Cooperative Highway Research Program, Report 197, Washington, DC, 1978.

40. G.R. Garner and R.C. Deen, "Elements of Median Design in Relation to Accident Occurrence," *Highway Research Record* 432, 1973, pp 1-11.

41. K. Moskowitz and W.E. Shaefer, "California Median Study: 1958," *Highway Research Record Bulletin* 266, 1960, pp 34-62.

42. S.C. Joshua and N.J. Garber, "Estimating Truck Accident Rate and Involvements Using Linear and Poisson Regression Models," *Transportation Planning and Technology*, Vol. 15, 1990, pp 41-58.

43. N.J. Garber and S.C. Joshua, *Traffic and Geometric Characteristics Affecting the Involvement of Large Trucks in Accidents*, Vol. II, Virginia Transportation Research Council, January, 1991, 50 pp.

44. M.E. Cassidy, *Fatal Accident Reporting System: Special Report on Heavy Trucks*, DOT-HS-803-259, NHTSA, U.S. Department of Transportation, Washington, DC, 1978.

45. G.R. Vallette, H.W. McGee, J.H. Sanders, and D.J. Enger, *The Effect of Truck Size and Weight on Highway Crashes, Vol. III: Accident Experience of Large Trucks*, FHWA/RD-80/137, Federal Highway Administration, Washington, DC, 1981.

46. Federal Highway Administration, *Highway Performance Monitoring System Field Manual*, Federal Highway Administration, Washington, DC, 1987.

47. S.-P. Miaou, P.S. Hu, A.K. Rathi, S.C. Davis, and T. Wright, *Development of Relationship Between Truck Accidents and Geometric Design: Phase I*, Draft Technical Memorandum No. 1, Oak Ridge National Laboratory, Oak Ridge, TN, March 1991.

48. Y. Mohamedshah, *Truck Accidents in Relation to Geometry of Highway: A Preliminary Report*, Internal Memorandum, Highway Safety Research Division, Federal Highway Administration, Washington, DC, September 1990.

49. F.M. Council and E.G. Hamilton, *Guidebook for the Utah State Data Files*, FHWA-RD-90-063, Federal Highway Administration, Washington, DC, January 1990.

50. F.M. Council and C.D. Williams, *Guidebook for the Illinois State Data Files*, FHWA-RD-90-080, Federal Highway Administration, Washington, DC, February 1990.

51. S.-P. Miaou, P.S. Hu, T. Wright, A.K. Rathi, and S.C. Davis, "Relationship Between Truck Accidents and Highway Geometric Design: A Poisson Regression Approach," paper presented at the 71st *Transportation Research Board Annual Meeting*, January 1991, and to be published in *Transportation Research Record*.

52. American Association of State Highway and Transportation Officials, *A Policy on Geometric Design of Highways and Streets*, Washington, DC, 1990.

53. C. Zegeer, R. Stewart, D. Reinfurt, F. Council, T. Neuman, E. Hamilton, T. Miller, and W. Hunter, *Cost Effective Geometric Improvements for Safety Upgrading of Horizontal Curves*, FHWA-RD-90-021, Federal Highway Administration, Washington, DC, May 1990, 227 pp.

54. G.E. Judge, W.E. Griffiths, R.C. Hill, and T.C. Lee, *The Theory and Practice of Econometrics*, John Wiley, New York, 1980, 793 pp.

55. Transportation Research Board, *Data Requirements for Monitoring Truck Safety*, Special Report 228, Transportation Research Council, Washington, DC, 1990.

56. Federal Highway Administration, *Traffic Monitoring Guide*, Federal Highway Administration, Washington, DC, 1985.

57. R.J. Carroll and D. Ruppert, *Transformation and Weighting in Regression*, Chapman and Hall, New York, 1988, 249 pp.

58. P. McCullagh and J.A. Nelder, *Generalized Linear Models*, Chapman and Hall, London, 1983, 261 pp.

59. D.R. Cox, "Some Remarks on Overdispersion," *Biometrika*, Vol. 70, No. 1, 1983, pp 269-274.

60. C. Dean and J.F. Lawless, "Tests for Detecting Overdispersion in Poisson Regression Models," *Journal of the American Statistical Association*, Vol. 84, No. 406, June 1989, pp 467-472.

61. D.R. Cox and P.A.W. Lewis, *The Statistical Analysis of Series of Events*, Chapman and Hall, London, 1966, 285 pp.

62. P.K. Andersen and R. Gill, "Cox's Regression Model for Counting Processes: A Large Sample Study," *The Annals of Statistics*, 10, 1982, pp 1100-1120.

63. E.L. Frome, D.L. Cragle, and R.W. McLain, "Poisson Regression Analysis of the Mortality among a Cohort of World War II Nuclear Industry Workers," *Radiation Research*, Vol. 123, 1990, pp 138-152.

64. A.C. Cameron and P.K. Trivedi, "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests," *Journal of Applied Econometrics*, Vol. 1, 1986, pp 29-53.

65. A.C. Cameron and P.K. Trivedi, "Regression-Based Tests for Overdispersion in the Poisson Model," *Journal of Econometrics*, Vol. 46, 1990, pp 347-364.

66. R.W.M Wedderburn, "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," *Biometrika*, Vol. 54, 1974, pp 439-447.

67. S.-P. Miaou and H. Lum, "Modeling Vehicle Accidents and Highway Geometric Design Relationships," *Working Paper, Oak Ridge National Laboratory*, Oak Ridge, TN 37831-6366, June 1992.

68. W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vettering, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1986, 818 pp.

69. J.S. Cramer, *Econometric Applications of Maximum Likelihood Methods*, Cambridge University press, New York, 1986.

70. H. Bozdogan, "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions," *Psychometrika*, Vol. 52, No. 3, 1987, pp 345-370.

71. A. Agresti, *Categorical Data Analysis*, John Wiley & Sons, New York, 1990, 558 pp.

72. M.E. Hallenbeck and L.A. Bowman, *Development of a Statewide Traffic Counting Program Based on the Highway Performance Monitoring System*, Federal Highway Administration, Washington, DC, March 1984.

73. Kansas State Department of Transportation, *Accuracy of the Streeter-Richardson TrafiComp System Used as a Vehicle Classifier*, February 10, 1989.

74. J.F. Lawless, "Negative Binomial and Mixed Poisson Regression," *The Canadian Journal of Statistics*, Vol. 15, No. 3, 1987, pp 209-225.

75. L.-F. Lee, "Specification Test for Poisson Regression Models," *International Economic Review*, Vol. 27, No. 3, 1986, pp 689-706.

76. B.J. Collings and B.H. Margolin, "Testing Goodness of Fit for the Poisson Assumption When Observations Are Not Identically Distributed," *Journal of the American Statistical Association*, Vol. 80, No. 390, June 1985, pp 411-418.

77. D.F. Signorini, "Sample Size for Poisson Regression," *Biometrika*, Vol. 78, No. 2, 1991, pp 446-450.

*U.S. G.P.O.:1993-343-273:80049

HSR-20/6-93(650)QE